Tour Into Painting : System Design for Virtual Exhibition of Chinese Hand-Scroll Painting

Yang Lu

THESE DE DOCTORAT DE

L'ÉCOLE CENTRALE DE NANTES

ECOLE DOCTORALE N° 602
Sciences pour l'Ingénieur
Spécialité : Architecture et Etudes Urbaines

Par
Yang LU

TOUR INTO PAINTING

System Design for Virtual Exhibition of Chinese Hand-Scroll Painting

Thèse présentée et soutenue à ENSA Nantes le 12 juillet 2022
Unité de recherche : UMR 1563, Laboratoire Ambiances Architectures Urbanités (AAU)

Rapporteurs avant soutenance :

Yun GAO — Senior lecturer, University of Huddersfield, U.K.
Pascal TERRACOL — Professeur, École Nationale Supérieure d'Architecture Paris Val de Seine

Composition du Jury :

Président : Pascal TERRACOL — Professeur, École Nationale Supérieure d'Architecture Paris Val de Seine
Examinateurs : Claus DIESSENBACHER — Professeur, Anhalt University of Applied Sciences, Dessau, Germany
Sang Bum PARK — Associate Researcher, Florida Agricultural and Mechanical University, U.S.A.
Dir. de thèse : Laurent LESCOP — Professeur des universités, École Nationale Supérieure d'Architecture de Nantes
Co-dir. de thèse : Jasper CEPL — Professeur, Bauhaus University, Weimar, Germany
Co-encadrant : Pascal JOANNE — Maître de conférences, École Nationale Supérieure d'Architecture de Nantes

Invité
Johannes KALVELAGE — Ingénieur de recherche, Anhalt University of Applied Sciences, Dessau, Germany

TOUR INTO PAINTING

System Design for Virtual Exhibition of Chinese Hand-Scroll Painting

A thesis submitted for the degree of
Doctor of Philosophy

LU Yang

AAU-CRENAU, UMR 1563 CNRS
SPI, Ecole Centrale de Nantes
FB3, Anhalt University of Applied Science

May 2022

Résumé

La peinture chinoise à rouleau (CHSP) est une forme de peinture chinoise typique. La manière correcte de regarder une CHSP est de la faire défiler à la main. Regarder une CHSP est considéré comme une expérience fantastique de visite dans son monde diégétique. Dans les expositions actuelles de CHSP, ces points ne sont pas bien représentés. Premièrement, la plupart des CHSP sont présentés dans des boîtes en verre. Les spectateurs ne peuvent pas les manipuler en les faisant défiler. Deuxièmement, l'expérience de la visite d'une peinture n'est interprétée que par de simples annotations. Les spectateurs ne peuvent pas acquérir une expérience complète à partir de ces annotations.

Cette recherche a exploré l'utilisation de la réalité virtuelle (RV) pour combler ces lacunes. L'objectif de cette étude est de développer un système d'exposition basé sur la RV, qui peut simuler de manière synchrone l'expérience de visualisation de l'ancien spectateur dans le monde réel et l'expérience de visite dans le monde diégétique.

Pour atteindre cet objectif, une étude en quatre étapes à été menée. La première est l'étude de la CHSP et de son processus de perception esthétique. Dans cette étude, les caractéristiques de la CHSP en termes de forme, d'image et de manipulation ont été résumées. Sur la base du processus esthétique général, la forme spatiale et la transition de la scène diégétique, ainsi que la position, la direction et le mouvement du spectateur dans la scène diégétique ont été déduits. La seconde est l'analyse de la structure visuelle d'une CHSP. Dans cette analyse, l'image d'une CHSP a été décomposée en trois couches, puis leurs structures visuelles et leurs méthodes de formation ont été analysées. Sur cette base, la corrélation géométrique entre la structure visuelle d'une CHSP et la structure spatiale de la scène diégétique a été établie, ainsi que la corrélation géométrique entre la structure visuelle et l'orientation et la position du point de vue diégétique. En outre, l'effet de la structure visuelle d'une CHSP à différents stades de développement sur le mouvement du point de vue diégétique et la variation de la profondeur dans la scène diégétique ont également été élaborés. La troisième est le développement d'un algorithme pour simuler le mouvement diégétique et la scène diégétique. Cet algorithme se compose de trois modules. Le Module 1 calcule la direction du rayon de projection passant par tous les points du CHSP, le Module 2 calcule la direction et la position du point de vue diégétique, et le Module 3 construire de la scène diégétique. Le quatrième étapes est l'implémentation du programme permettant de corréler l'expérience visuelle dans la scène réelle et l'expérience touristique dans la scène diégétique. Dans ce programme, un espace virtuel imbriqué a été conçu, dans lequel une scène d'observation contenant une CHSP virtuel et une plate-forme d'observation a été placée dans la scène diégétique. Ensuite, un programme de corrélation a été développé pour synchroniser la position du point focal du spectateur sur la CHSP virtuelle avec la position et l'orientation de la plate-forme de visualisation dans la scène diégétique, ainsi que l'image dans le cadre déroulé avec la scène diégétique.

Le système a été présenté comme une application de RV comprenant une interactive CHSP qui peut être manipulé selon les principes originaux. De manière synchrone, l'utilisateur de la

RV sera déplacé en fonction de son point de focalisation sur la CHSP virtuelle, puis un monde diégétique englobant et aléatoire sera construit.

Abstract

The Chinese Hand-Scroll Painting (CHSP) is a typical Chinese painting form. The proper way to view a CHSP is to scroll it by hand. Watching a CHSP is considered a fantasy experience of touring in its diegetic world. In current exhibitions of CHSPs, these points are not well represented. First, most of CHSPs are presented in glass boxes. Viewers can not manipulate it in a scrolling manner. Second, the experience of touring in the painting is only interpreted by simple annotations. Viewers can not gain a full experience based on them.

This research explored to use of Virtual Reality (VR) to implement these shortages. The goal of this study is to develop a VR-based exhibition system, which can synchronously simulate the ancient viewer's viewing experience in the real world and the touring experience in the diegetic world.

To achieve this purpose, four-phased studies were conducted. First is the study of CHSP and its aesthetic perception process. In this study, the characteristics of CHSP in terms of form, image, and manipulation were summarized. Based on the general aesthetic process, the spatial shape and transition of the diegetic scene, as well as the position, direction, and movement of the viewer in the diegetic scene were deduced. The second is the analysis of the visual structure of CHSP. In this analysis, the image of CHSP was deconstructed into three layers, and then their visual structures and formation methods were analyzed. Based on this, the geometric correlation between the visual structure of CHSP and the spatial structure of the diegetic scene was established, as well as the geometric correlation between the visual structure and orientation and position of the diegetic viewpoint. Besides, the effect of the visual structure of CHSP at different developmental stages on the movement of diegetic viewpoint and depth variation in the diegetic scene were also elaborated. The third is the algorithm development for simulating the diegetic movement and the diegetic scene. This algorithm consists of three modules: Module 1 for calculating the direction of the projection ray passing through all points on the CHSP, Module 2 for calculating the direction and position of the diegetic viewpoint, Module 3 for constructing the diegetic scene. Fourth is the program implementation for correlating the viewing experience in the real scene and the touring experience in the diegetic scene. In this program, a nested virtual space was designed, in which a viewing scene that contains a virtual CHSP and viewing platform was placed in the diegetic scene. Then, a correlation program was developed to synchronize the position of the viewer's focus point on the virtual CHSP with the position and orientation of the viewing platform in the diegetic scene and the image in the unrolled frame with the diegetic scene.

The system was presented as a VR application that includes an interactive CHSP that can be manipulated according to the original principles. Synchronously, the VR user will be moved according to his focus point on the virtual CHSP, and then, an encompassing, random diegetic world will be built.

Acknowledgement

This research project was established in the collaboration between Anhalt University of Applied Sciences, École Nationale Supérieure d'Architecture de Nantes, and École Centrale de Nantes, and was financed by a scholarship from Anhalt University of Applied Sciences. I am proud to be the first Ph.D. graduate from this collaboration, and I sincerely thank these three universities for all the support they have provided me.

This research could not have been completed without the generous support from my thesis committee. First of all, I would like to thank my thesis advisor, Prof. Laurent Lescop, for his invaluable supervision, encouragement, and generous assistance from the inception to the completion of my research. It was his encouragement that helped me to make a firm decision to start my academic research. I thank him for his patience in guiding and inspiring me at all stages of my research. Those research maps he drew for me in every meeting were the best proof of this. Also, thanks to his generosity in sharing and discussing all kinds of information with me, taking me to conferences and exhibitions, and showing me the prospect of combining classical art and VR technology.

I would also like to thank Prof. Pascal Terracol of École Nationale Supérieure d'Architecture Paris-Val de Seine and Prof. Yun Gao of Huddersfield University for reading my thesis and giving me enlightening comments. Meanwhile, I would like to express my thanks to Prof. Pascal Joanne of ENSA Nantes and Dr. Sang Bum Park of Florida A&M University for examining my thesis.

I am very grateful to Prof. Jasper Cepl of Bauhaus University, Prof. Claus Diessenbacher of Anhalt University of Applied Sciences, and Prof. Jacque Gilbert of Nantes University for their annual evaluations of the phased studies.

I owe a huge debt of gratitude to Prof. Johannes Kalvelage of Anhalt University of Applied Sciences. He supervised my master's thesis and launched my academic career. During my doctoral research, he also provided valuable academic advice and helped me coordinate administrative matters.

I am so appreciative of the laboratory AAU-CRENAU for providing me with excellent working facilities. Meanwhile, I also want to thank my colleagues in the lab. They added countless pleasures to my academic life. They also participated in my VR experiments and provided me with valuable suggestions.

Last, but not least, I owe hugely to my family for their constant love and understanding. Without their support and encouragement, I could not have finished this research.

Contents

Chapter 1

Introduction

1.1 Background

1.1.1 Virtual exhibition of painting

Virtual Reality (VR) techniques have been widely applied in the exhibition of paintings, which creates a new form of virtual exhibition with immersive features. The most popular one is the "Virtual Tour in the Museum" of Google Arts & Culture[1]. The developers reconstructed famous galleries in cyberspace with the technical support of *Google Street View*. By wearing a VR headset or even low-profile cardboard VR glasses, the visitor can easily make a virtual tour of galleries without restrictions on time and space. The VR-based exhibition presents a huge advantage in improving the accessibility of artworks. Therefore, it soon became a common auxiliary means for ordinary exhibitions in famous galleries and museums around the world and even became an alternative for an actual exhibition during the COVID-19 period when all galleries and museums were closed (Burek, 2020).

The development of the VR-based exhibition of paintings was influenced by the concept of New Museology. This concept claims that the museum's role should be shifted from collecting artworks (Mairesse and Desvallées, 2010) to increasing access and educating the visitors (Vergo, 1997; McCall and Gray, 2014). This claim puts two demands on VR-based exhibitions: promoting a more complete viewing experience and providing a more intuitive interpretation. Accordingly, two types of VR-based painting exhibitions have been developed.

One is the immersive representation of the exhibition space. Typical examples are the *Petit Galerie*[2] of the Louvre [Figure 1.1.1-left] and the *Digital Dunhuang*[3] of the Dunhuang Mogao Grottoes [Figure 1.1.1-right]. This type of exhibition usually contains several virtual exhibition halls, which are digital replicas of the real ones. They are reconstructed by 3D modeling or photogrammetry techniques or, more commonly, are presented as spheres whose inner surface is the panoramic projection of a real exhibition hall. The visitor can enter these virtual showrooms with the assistance of a VR headset and freely perform regular actions in them. For example, watching around from a fixed standpoint, or moving around in a small area. In the virtual exhibition hall, the paintings remain in their original state of display and are accompanied by interactive panels. When the viewer activates this panel by clicking a button on the VR joystick or glasses, the painting will be displayed as a high-resolution image with corresponding textual

[1] https://artsandculture.google.com/project/360-videos
[2] https://petitegalerie.louvre.fr
[3] https://www.e-dunhuang.com

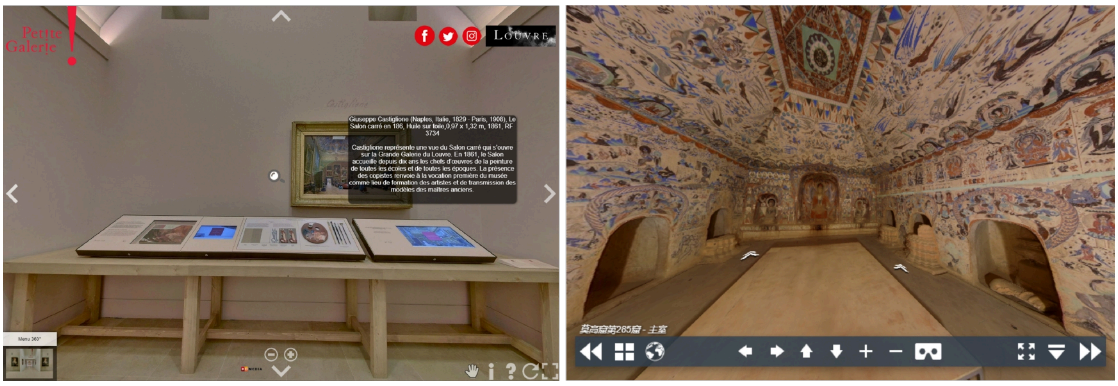

Figure 1.1.1: First type of virtual exhibition

or audio annotations.

Since this virtual exhibition reconstructs the display scene and simulates the viewing method, this virtual exhibition ultimately achieves an immersive representation of the viewer's viewing experience in the real world. This immersive representation has advantages in exhibiting those paintings that must be experienced *in situ*. Such as the murals drawn on the inner faces of Dunhuang Gottos. [Figure 1.1.1-right] These murals do not exist as isolated images, they are related in content and form an inseparable total artwork with the atmosphere of the cave.[4] (Wu, 2018). One of the requirements of viewing such paintings is to experience them as a whole. From this point of view, this virtual exhibition, which reconstructs the original display scene, is able to provide a more comprehensive experience of the paintings, in contrast to displaying them isolated in a glass cabinet. At the same time, viewing such murals is a form of Buddhist practice that requires following a specific ritual. Ancient practitioners sat on or around a pedestal in the center of the grotto and then viewed multiple murals around them simultaneously through a mirror array that surrounded them. (Wang, 2004) From this perspective, this virtual exhibition, which allows the visitor to freely move in the virtual exhibition hall, has the ability to allow them to experience the paintings in the way it was originally intended (although it has not yet been realized in the current project).

The other is the intuitive interpretation of the painting's diegetic space[5]. Typical examples are *Zaha Hadid's Paintings: A Virtual Reality Experience* of ZHVR Group[6] [Figure 1.1.2-top left], *Dreams of Dalí* of The Dalí Museum (St. Petersburg)[7] [Figure 1.1.2-top right] and *Mona Lisa: Beyond the Glass* of the Louvre[8] [Figure 1.1.2-bottom]. The most difference from the previous type of virtual exhibition is that this virtual exhibition presents not only the painting itself but mainly the diegetic scene of the painting. In most cases, this diegetic scene is usually constructed by VR developers based on their own imagination of the painting space and using elements from the painting. While in some more serious projects it is constructed strictly according

[4]Total artwork (Gesamtkunstwerk) comes from Wilhelm Richard Wagner's essay of *The Artwork of the Future (Das Kunstwerk der Zukunft)* He broke down the boundaries between artistic fields, integrating painting, architecture, music, performance, and scenography in the theater to form a comprehensive presentation of various arts. In the 20th century, artists introduced this concept to other artistic fields to define the complex interactions between different arts within a perceptual place.

[5]"Diegetic space" is a concept from cinematography. It refers to the world in the narration. Here I used it to refer to the represented space inside the painting

[6]https://www.zhvrgroup.com/zaha-hadid-paintings-virtual-reality-experiences

[7]https://thedali.org/exhibit/dreams-of-dali-in-virtual-reality/

[8]https://www.louvre.fr/en/what-s-on/life-at-the-museum/the-mona-lisa-in-virtual-reality-in-your-own-home

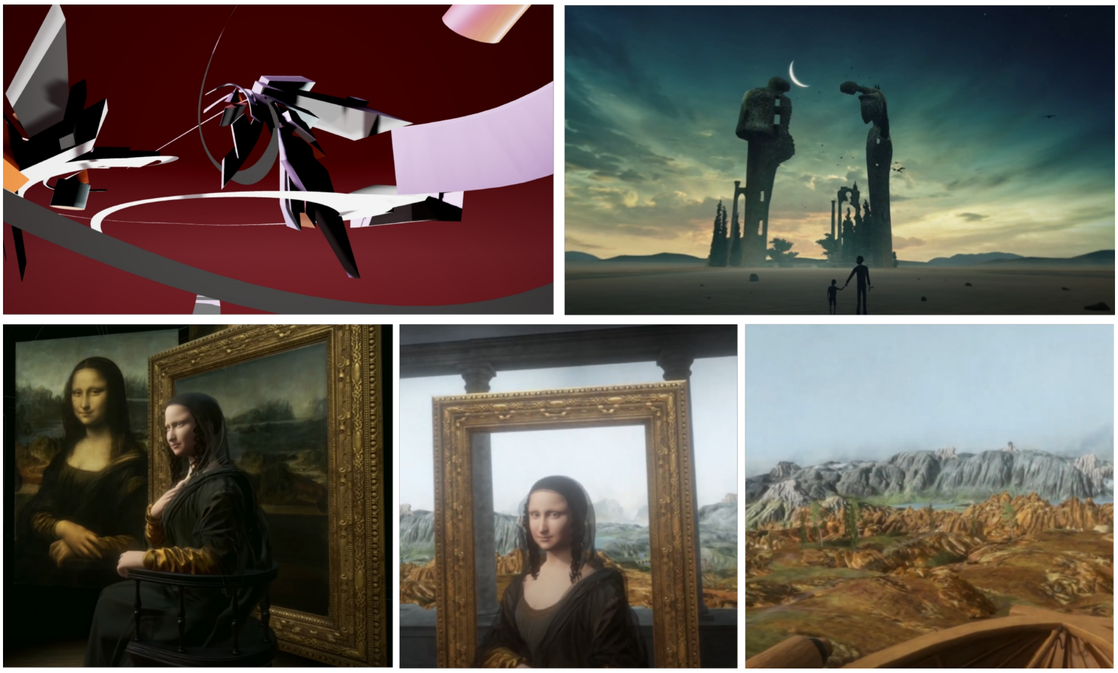

Figure 1.1.2: Second type of virtual exhibition

to the relevant research. For example, in the project of *Mona Lisa: Beyond the Glass*[9], the reconstruction of the mysterious landscape behind Mona Lisa is constructed based on rigorous investigations.[10] [Figure 1.1.2-bottom] In this sense, this diegetic scene is a synthesis and visual interpretation of the research results of the painting.

In addition, this type of VR-based exhibition has a considerable advantage in creating unconventional experiences. Typically, the viewer is first placed in a virtual gallery where a virtual painting is displayed, and then gradually crosses the canvas and finally becomes fully immersed in the painting's diegetic scene. In this scene, the visitor is not only able to perform realistic movements, such as spinning his head and walking, but also non-realistic movements, such as floating in Zaha's deconstructionist space or flying in the fantasy world of a Dali painting.

With the development of VR-based exhibitions, these two types are also starting to be integrated, thus achieving the unity of presentation and interpretation.

1.1.2 Tour into Chinese Hand-Scroll Painting

Touring in the diegetic world of painting is not a novel idea of modern people, ancient Chinese believed that watching a painting was not only a purely visual experience in the real world but more than as a bodily experience in the diegetic world of the painting (Li, 2009). One of the most influential concepts is *Wo You* 卧游(recumbent tour) proposed by the painting theorist Zong Bing 宗炳(375 - 443). According to the *Song Book* 宋书, Zong loved to travel when he was young. When he grew old and needed to return to Jiangling to convalesce, he had the landscape

[9]https://www.louvre.fr/en/what-s-on/life-at-the-museum/mona-lisa-beyond-the-glass-the-louvre-s-first-virtual-reality-experience

[10]https://www.washingtonpost.com/archive/lifestyle/1995/12/14/the-mystery-behind-the-mona-lisa/cad1a4af-f9e3-4463-a609-6cdac7923e70/

Figure 1.1.3: Reading in an Open Hall. Zhao Bosu (1124 - 1182). Taipei Palace Museum, Taipei

paintings set up indoors, replacing travel with viewing them.

> (Zong) said with a sign, "I'm old and sick, so touring famous mountains and rivers are now quite beyond me. What I should do is unleash my soul and look inwardly to seek the Dao 道[11]. Then I watch a landscape painting as if I were actually traveling there, even when I am still lying in bed."[12] (Shen, 1974)

In his theoretical work *Preface to Painting Landscapes* 画山水序, Zong further refined and theorized this concept through the "Li"理[13]. Zong indicated that the landscape painting should represent the "Li" of the landscape.

> The painter should first experience the landscape and comprehend the "Li" of nature. Then he should put the spirit of the landscape in the form and integrate his emotions into the painting so that the "Li" will also enter the landscape painting work. ...... If the painting is really like this, even a square inch of it can express the charm of mountains and rivers and the spirit of nature.[14]

And when the painting represents the "Li" of the landscape, then viewing such a landscape painting is equal to, or even better than, visiting a real landscape.

> Then, (as I) sit there leisurely, drinking wine and playing the Qin 琴(musical instrument), I unfolded a scroll to explore the scenery in the painting. What I saw, at this moment, was the jungle of the sky, or the wild scenery of no one. There are both cliffs and clouds and forests.[15]

Zong described the experience of disengaging from reality and losing himself in the diegetic world of the painting, while he was watching a painting in the bed. During this experience, the experience paradigm of watching a landscape painting transcends from contemplating a landscape image to being in a diegetic landscape (Law, 2011).

[11] Or spelled as Tao, the nature of the landscape

[12] Original text: （宗炳）好山水，爱远游，西陟荆、巫，南登衡岳，因结宇衡山，欲怀尚平之志，有疾还江陵，叹曰:"老疾俱至，名山恐难遍睹，唯澄怀观道，卧以游之。"

[13] Same as Dao, the nature of the landscape.

[14] Original text: 夫以应目会心为理者，类之成巧，则目亦同应，心亦俱会。应会感神，神超理得。虽复虚求幽岩何以加焉？又神本亡端，栖形感类，理入影迹，诚能妙写，亦城尽矣。

[15] Original text: 于是闲居理气，拂觞鸣琴，披图幽对，坐究四荒，不违天励之，独应无人之野。峰岫嶷，云林森眇。

Figure 1.1.4: A Thousand Li of Rivers and Mountains. Wang Ximeng (1096-1119). Palace Museum, Beijing

Zong's fantasy experience of being in the painting has been recognized and resonated with later generations of artists and theorists. They construct an aesthetic system of classical painting centered around the concept of "tour in the painting". The painter also aimed to create paintings that could be "being in" from then on. In his article of *Lin Quan Gao Zhi* 林泉高致, Guo Xi's 郭熙(c.1020 - c.1090) once defined that the excellent landscape painting should allow the viewer to be in.

> There is a precise way to say that a natural landscape can be strolled in, can be climbed and looked at from afar, can be suitable for a relaxing trip, and can also be enough for a comfortable residence. If a landscape painting can achieve "feasible, watchable, livable, and tourable", it is an "excellent piece".[16]

Accordingly, the experience of "being in" the painting's diegetic world became an integral part of the complete viewing experience of the painting. Just as Guo further described the experience of watching a landscape painting,

> "even without leaving the pavilion, I can enjoy the clarity of the springs and ravines, the singing of birds and apes, and the sparkling of mountains and water."[17] [Figure 1.1.3]

It is to be noted that the experience of "being in the painting" discussed here is a general bodily experience, which corresponds to a conventional single painting. However, due to the different forms and interactions of paintings, there are various peculiarities in the ways of being in.

Chinese Hand-Scroll Painting (CHSP) is a special form of classical painting that is painted or mounted on long but narrow rice paper or silk. [Figure 1.1.4] As this form is suitable for composing multiple images in the horizontal direction, the subjects of such paintings in early times are mostly related to narratives, such as mythological and religious stories, historical events, etc. However, from the 4th century onwards, the landscape became the mainstream subject matter. CHSP was used by the artist to record what he saw during his journey. In making such CHSP, the artist followed an editing way like the montage in cinematography. He captured images of different scenes from the same time, or the same scene from different times, or different scenes from different times and combined them horizontally in the same frame. The scroll painting is therefore seen as an edited landscape film.

[16]Original text: 世之笃论，谓山水有可行者，有可望者，有可游者，有可居者。画凡至此，皆入妙品。

[17]Original text: 不下堂筵, 坐穷泉壑；猿声鸟啼, 依约在耳；山光水色, 滉夺目。

Figure 1.2.1: CHSP exhibition in Victoria and Albert Museum, London

The way of viewing a painting is usually matched to its form. The long form of CHSP prevents it from being fully expanded at one time, thus requiring the viewer to adopt a scrolling manner. The viewer needs to unroll a certain part each time to watch, and then roll up the part already seen and open a new one. This operation allows the images in the CHSP to be displayed progressively, thus ensuring a reasonable narrative of the story. It is for this reason that the CHSP has been used as a perfect image narrative vehicle for a long time.

The image characteristics and viewing methods of CHSP give the experience of CHSP a spatial narrative feature. If viewing a painting with a single image corresponds to the experience of "being in" a static diegetic scene, then, the continuous switch of the displayed images in the CHSP means that the continuous transition of their corresponding diegetic scenes. The audience is thus given a dynamic experience, traveling from one scene to another. Following the idea of metaphorizing the painting process of the CHSP as a montage, then the viewing process is the screening of a film. While, the difference is, when to screening a film, the audience is passive, and the narrative is controlled by the director. But when watching a CHSP, the viewer is active and he can control the speed and even reverse the narrative direction by rolling back the scroll.

1.2 Motivation

As one of the most extant classical Chinese paintings (Xue, 1993), CHSPs are not only collected and exhibited in the galleries and museums of China but also in those of European countries, such as the Musee Guimet in France, the British Museum in the UK, etc. In these galleries and museums, the most common exhibition way is to unroll the CHSPs completely and place them in glass cabinets. The visitors can only watch them from a distance and understand them with the assistance of brief text or audio annotations. [Figure 1.2.1]

This presentation protects CHSPs to the greatest extent possible, as they become fragile after being stored for hundreds of years. However, it also causes two shortages. First, the visitor cannot manipulate the CHSP in a scrolling manner. As introduced before, this manner ensures that the images in CHSP appear in a progressive manner, which is the basis for a proper narrative. When CHSP is fully opened, it will cause the same scene or character at different times to appear simultaneously, which may cause a Spatio-temporal paradox. For example, the painting of *The Night Entertainments of Han Xizai* 韩熙载夜宴图. [Figure 1.2.2] If this scroll is fully unfolded, Han Xizai, the person in the box will appear in different rooms simultaneously, which leads the viewer to mistakenly believe that these characters are different persons. But if

Figure 1.2.2: The Night Entertainments of Han Xizai, Gu Hongzhong, Palace Museum, Beijing

the painting is scrolled, this confusion will not happen. Second, simple annotations can not help the viewer to gain the experience of touring into painting in a short period of time. As Zong Bing stated, to realize the recumbent tour, a preparatory activity of "unleashing soul and looking inwardly to seek the Dao" is needed. According to Clunas (1985), this activity is essentially a practice of Taoist introspection, which is impossible to achieve with a plain explanation. Based on the above two shortages, it can be judged that the current exhibition method cannot provide a complete viewing experience of the CHSP.

However, as discussed in 1.1.1, VR technology has great advantages in simulating the realistic experience in the exhibition hall and the non-realistic experience in the diegetic space of the painting. Therefore, it can be a solution to complement these shortages in the exhibition of the CHSP by simulating the manipulating experience, as well as the touring experience. Based on this idea, this research aims to develop a VR-based exhibition system for CHSPs. This system provides a virtual scroll and allows the VR user to manipulate it in an original way. While the VR user is watching this virtual scroll, this system will build a corresponding diegetic scene in real-time and direct the VR user on a virtual tour of this scene.

1.3 Methodology

To achieve the purpose above, two experiences need to be simulated in a virtual environment: the experience of the viewer watching CHSP in the real world and the experience of the viewer touring in the virtual world.

As for the simulation of the former experience, the CHSP is an objective entity, and watching CHSP is definite action. Therefore, this simulation can be achieved by conventional simulation methods. That is, create a virtual CHSP by using 3D modeling techniques and then develop the interactive programs to simulate interactions, such as scrolling and unfolding, etc.

The challenge of the research is the simulation of the latter experience. The experience of "tour in the painting" is purely a subjective fantasy of the ancient viewer. There does not exist a diegetic scene with a definite spatial form in the real world. At the same time, the ancient artist's description of "tour in the painting" does not contain a definitive description of the movement, like path, or method. This means that the simulation of the latter experience is not a digital representation of an existing reality in the usual sense, but is closer to the realization of a vision in architectural or scenic design, based on a rendering. Therefore, the simulation of the latter experience requires a set of development methods similar to scene construction and path guidance in spatial design. Therefore, this research integrates the methods of spatial design and formal analysis of images into VR development to realize the simulation of the viewer's experience in the diegetic scene. The entire research is divided into four parts: spatialization of the diegetic experience, geometrization of diegetic scene and movement, the direction of diegetic movement and the construction of diegetic scene, synchronization of the real and diegetic experiences.

1.3.1 Spatialization

In spatial design, when we want to represent an abstract concept or a vague fantasy into a concrete experience, we need to "spatialize" it. Specifically, it is to determine the form of the space being experienced and the way in which a man experiences that space. One principle of this process is that the composition of the space and the way of movement must be set in such a way that the final spatial experience is consistent with the subject matter that is intended to be expressed. Since the experience of touring in a painting is an illusion based on visual perception, which is sensual and ambiguous, the representation of it also requires spatialization.

In current projects, the most common means of spatialization is to build the diegetic scene of the painting and set a preset movement path based on the VR maker's personal perception and imagination of the painting. A typical example is the VR project of *Walk into the painting of Autumn Colors of Que and Hua*[18] created by the Palace Museum in Taipei. In this project, the shape of the diegetic scene and the way and path of the diegetic movement are all designed by the VR maker. This approach is the easiest way to realize spatialization, but it is always accompanied by a controversy: is the experience, that the VR user gets, set by the painter or by the VR maker? This issue is irrelevant in entertainment-oriented VR projects, as they aim to provide a joyful immersive experience. However, if we want to use the VR project as an interpretation of the painting, then we must ensure that the diegetic scene and diegetic movement match the ancient painter's intention. This means that in the development of VR projects for such serious purposes, we need to restore the experience of the ancient Chinese to the greatest extent possible.

For this reason, I abandoned the spatialization method based on personal perception and imagination, and chose to infer the corresponding diegetic scene and diegetic movement of the CHSP according to the generation law of "touring in the painting". First, I organized the aesthetic experience process of an ordinary painting in the context of classical Chinese aesthetics. In this process, I summarized two transformation processes. One is the transformation from the 2D painting to the 3D diegetic scene. The other is the transformation from watching the painting to being in the diegetic scene. On this basis, I combined the classical Chinese concept of space and the ancient projection method to infer the spatial form of the diegetic scene of the general painting, as well as the position and orientation of the viewer in it. Second, I analyzed the specificity of CHSP compared to the general painting with a regular framed form. In the analysis, I refer to Wu's (1996) view to treat the CHSP as the collaboration between the image and the image bearer. The whole analysis revolves around the CHSP's form and encompasses the image and interaction of the CHSP. Finally, I combined the specificities of CHSP and the experience of "being in the regular framed painting" to infer the experience of "touring in the CHSP". In this inference, I treat the image in the unrolled frame as a regular painting, thus establishing a correspondence between this image and the diegetic scene and diegetic viewpoint. Based on these correspondences, I further combined the interactions of scrolling and moving viewing to reason the narrative way of the diegetic scenes and the movement way of the diegetic viewpoint corresponding to the CHSP.

1.3.2 Geometrization

The spatial shape of the diegetic scene and the path of the diegetic movement derived after spatialization is only sensory but not precise. Therefore, they cannot yet be directly used for the digital representation that requires definite geometric data. For this reason, we need to obtain the exact geometric information of the diegetic scene and diegetic movement from the CHSP.

[18]https://www.youtube.com/watch?v=MNuMH_OvptU

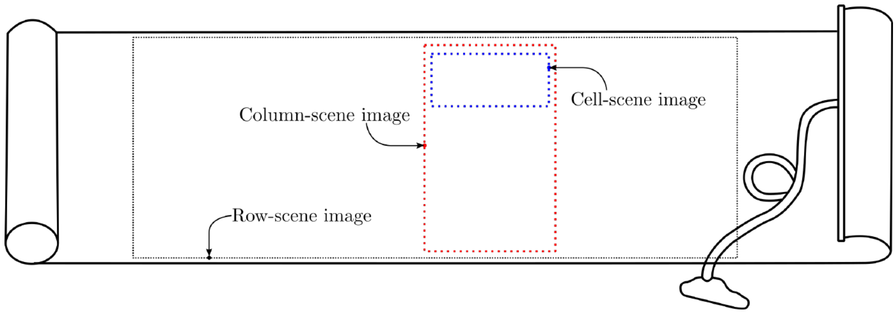

Figure 1.3.1: Three layers of CHSP's image

Visual perceptual studies (Arnheim, 1954, 1977; Kadar, 2008) and painting study (Loran, 2006) have demonstrated that the visual system's estimation of the position and orientation of the painter's viewpoint and the 3D position of objects are related to the visual structure of the painting. Visual structure is the structural relationship between forms in an image, such as perspective. It implies the transformation rules that the painter follows when converting a three-dimensional object into a two-dimensional image, such as the perspective transformation model. Both the geometric information of the painter's viewpoint and the represented scene are correlated to the transformational model of visual structure. Therefore, the visual structure of the painting, like perspective, was used as a calculation basis in the projects of painting-based 3D reconstruction (Horry, 1997; Saegusa, 2004; Sdegno, 2004; Furferi, 2014) and viewpoint estimation (Aubry, 2014). However, the problem is that there is not yet a well-developed transformation model of the visual structure of the CHSP. Although some computer scientists have developed a transformation model for the CHSP (Ma, 2011, 2012), this model can only yield approximate depth information, but not other necessary information, such as the exact position of the viewpoints. For this reason, a transformation model for the CHSP needs to be established.

To build this transformation model, the process of the painter converting the 3D scene into a CHSP needs to be geometrized. There are two factors that need to be considered in the geometrization process. One is the complexity of CHSP's visual structure, which is reflected in its multi-layered nature. According to Bachhofer's (1931) studies on ancient Buddhist murals, a basic process to form an image of a large scene is to use the image of small scenes as "space cells" to compose an image as the representation of a large scene, instead of projecting a large scene directly. This logic is embodied in three operations in the CHSP. Firstly, the painter projected images of small scenes with different depths. Secondly, the painter vertically combines these small scene images to form an image of the scene with a larger depth. Finally, the painter horizontally stitches these larger-depth scene images to form an image sequence. These operations made the CHSP a nesting visual structure with three layers. For analyzing convenience, I defined these three layers as, cell-scene image, column-scene image, and row-scene image, as shown in [Figure 1.3.1]. The cell-scene image is the result of the operation of projection. It refers to a spatial unit, in which all object images have a unified visual structure. The column-scene image is the result of the operation of vertical composition. It refers to a temporal unit. It is the image that the painter sees at a time. The row-scene image is the result of the operation of horizontal composition. It refers to the entire image narrative. On the basis of the above division, I first inferred the formation method of the cell-scene image by combining ancient painting theory and architectural graphical theory. Second, I derive the geometric correlation between its visual structure and the orientation and position of the diegetic viewpoint, and the geometric correlation between its visual structure and the 3D position of the object in the diegetic scene, based on its formation

Figure 1.3.2: Mural in the Cave 290 and the Cave 61 of Dunhuang Grottos

method.

The other factor to be considered is the multiplicity of the visual structure of CHSP, which is reflected in the way cell-scene images are combined in column-scene images and row-scene images in the CHSPs at different historical periods. Bachhofer (1931) also study of Buddhist murals also found that there was an evolution in the way ancient painters combined small scenes to form larger ones. This evolution started from the isolated juxtaposition of small scene images with clear boundaries and then evolved to the complete integration of small scene images with dissolved borders, just like the transformation from the visual structure of the mural in the Cave 290 [Figure 1.3.2-right] to the visual structure of Wutai Mountain mural in the Cave 61 [Figure 1.3.2-left]. According to Fong (1969), when a painter combines basic form elements or motifs and compositional patterns to create new effects or find solutions to new problems, he would inevitably create form relationships and visual structures characteristic of his own time. This means the same composition of scene images in the Dunhuang murals would also appear in the CHSPs during the same period. Therefore, we can generally divide the development of CHSPs according to the evolution of the compositions in the Dunhuang murals. Based on this, I divide the development period of CHSP into three stages according to the way cell-scene images are combined in CHSP.[19] Stage 1: before the early 11th century. The composition of cell-scene images in the CHSP at this stage is characterized by compartmental juxtaposition. Stage 2: late 11th-early 13th century. The composition of cell-scene images in the CHSP at this stage is characterized by partial overlapping. Stage 3: late 13th-14th century. The composition of cell-scene images in the CHSP at this stage is characterized by open fusion. Different compositions mean different switch ways of the cell-scene images when the viewer is watching the CHSP. And the switch would further affect the diegetic movement and the way the depth of the object changes in the diegetic scene. To summarize the patterns of change, I selected representative CHSPs from each stage for analysis. First, I analyzed the way that cell-scene images are composed in the column scene image and row scene image. Second, I combined the two related geometric correlations of cell-scene images and their compositions in the column-scene image and row-scene image to infer the way the viewer's diegetic viewpoint moves while he is watching the column-scene image and row-scene image, as well as the way the depth changes in the column-scene and row-scene.

1.3.3 Computation & Construction

After inferring spatialization and geometrization of the diegetic scene and diegetic movement, it is possible to use these results to develop an algorithm to construct the diegetic scene and direct

[19]This division also refers to Fong's (2003) division on the development period of classical Chinese paintings.

the diegetic movement in VR.

For this task, Chu and Tai (2001) propose an idea in their project of "tour into the CHSP". They first divided the CHSP into sub-scene images. Second, they calculated the position and orientation of the corresponding viewpoint and the 3D positions of the objects of each sub-scene image based on the transformation model of perspective. Finally, they separately rebuilt these sub-scenes and set the positions of these viewpoints as the camera's points. This approach achieves an effect of the multi-scene narrative, but it is not consistent with the spatial narrative derived in 1.1.2. First, the division they use makes the CHSP subjectively segmented into isolated image fragments, which makes it likely that scene images that were not originally in the same space-time are combined or split up from the same scene image. Second, the perspective transformation model they used does not match the visual structure of CHSP when calculating the position and orientation of the diegetic viewpoint and constructing the diegetic scene. This makes each sub-scene image correspond to only one fixed diegetic viewpoint and one isolated sub-scene. Therefore, when the user wants to move from one sub-scene to another, he needs to exit the current sub-scene first and secondly enter another sub-scene. Obviously, this experience is not consistent with the experience of continuous space represented by CHSP.

In order to make the simulated spatial experience conform to the CHSP representation, I designed a method that strictly follows the transformation model of the CHSP visual structure. The first step of this method is the splitting of CHSP. The geometrization points out that the CHSP uses the cell-scene image as the basic unit. Therefore, the splitting of sub-scene images in CHSP should be done in the cell-scene image. According to the definition of cell-scene image, it refers to a spatial unit, in which all object images have a unified visual structure. Each kind of visual structure corresponds to a cell-scene image. Thus, the difference in visual structure can be used as a basis for distinguishing sub-scenes. In the splitting process, I first identify different visual structures of cell-scene images based on the compositions in the column-scene image and row-scene image. Then, I separate the cell-scene images from the CHSP according to their different visual structures.

The second step is the calculation of the position and orientation of the diegetic viewpoint. During the calculation, I first calculated the direction of the projection ray passing through each point in the CHSP, based on its geometric association with the visual structure of the cell-scene image. Second, I calculated the position and orientation of the diegetic viewpoint corresponding to each point, based on its geometric correlation with the projection ray.

The third step is the construction of the diegetic scene. Unlike Chu's project where only the scene being represented in CHSP is reconstructed, the diegetic scene we need to build needs to be panoramic, thus satisfying the need for VR users to view freely in it. This means that the construction of the diegetic scene has two parts: reconstruction of the scene that is represented in the CHSP and construction of the rest scenes outside represented scenes that are not depicted in the CHSP.

According to the analysis of the manipulation method of CHSP in 1.1.2, CHSP can only be unrolled a part at a time. Therefore, the represented scene here is the scene corresponding to the image in the unrolled frame. I designed a "reverse engineering" approach, that is, to reconstruct the represented scene based on the reverse process of the CHSP's formation. It consists of three steps. First, I segmented each object image from CHSP. Second, I calculated the 3D position of each object according to the transformation model of the cell-scene image's visual structure. Third, I used the ground-contact of the object as the anchor to relocate each segmented object image to its 3D position. Last, I restored the 3D shape and size of each object based on the transformation model of the cell-scene image's visual structure.

The rest scene is not represented in CHSP, it is essentially a fantasy of the viewer. It is uncertain. Therefore, we need to build a scene with uncertainty. Here I use randomness to express uncertainty by designing a method of "random generation". This method is implemented in two steps: First, I constructed the basic spatial form of the rest scene based on the inference of the spatial form of the diegetic scene. Second, I created and arranged objects randomly in the rest scene.

1.3.4 Synchronization

After developing the computation and construction methods related to the diegetic scene, it is time to represent the diegetic experience of "touring in the CHSP" in the virtual environment.

According to the analysis in 1.1.2, this experience has two characteristics. First, touring experience in the diegetic world corresponds to viewing experience in the real world. The diegetic scene is imagined based on the image in the unrolled frame. At the same time, the viewer's diegetic viewpoint corresponds to his focus point, thus ensuring that the diegetic scene he sees is consistent with the orientation and position of the image presented in the unrolled frame. Second, the diegetic movement is a dynamic and viewer-directed action. The viewer needs to scroll the CHSP and move the focus point while watching it. These manipulations respectively make the image in the unfolded frame dynamically change and the position of the focus point change. Since diegetic experience corresponds to the viewing experience, the diegetic scene would change and the viewer's diegetic viewpoint would move synchronously. From this point, the viewer's experience in the diegetic scene is a dynamic movement controlled by himself. These two features are fundamental to the aesthetic experience of the CHSP, which needs to be interpreted in the exhibition system.

For this purpose, I designed a nested virtual space, integrating a viewing scene into an initial diegetic scene. This viewing scene consists of a platform and a virtual scroll, where the VR can stand on and manipulate a virtual scroll. The initial diegetic scene is the 3D model of the represented scene of the entire CHSP. Second, I developed an interactive program to synchronize the VR user's interactions in the viewing scene and his movement in the diegetic scene. The program will detect the position of his focus point on the virtual scroll, while the VR user watches the virtual scroll. Then, it will calculate the direction and position of the diegetic viewpoint corresponding to this focus point. And lastly, it will adjust the platform's orientation and position to the calculated values and build up the diegetic scene in real-time. Third, I developed a construction program to synchronize the image in the unrolled frame and the diegetic scene. The program will detect the four corners of the unrolled frame and set up the area of the represented scene, and then adjust the displayed part of the initial scene and construct the rest scene in real-time by following the methods developed in **Computation & Construction**. Last, I integrated these programs into the VR virtual exhibition system and execute them in real-time as the VR user interacts with the virtual CHSP.

1.4 Mapping of thesis

The whole thesis is divided into four parts corresponding to the four main steps in **Methodology**.

The first part is the **Spatialization**, which contains two chapters. Chapter 2 summarizes the aesthetic process of being in classical Chinese painting, and infers the spatial morphology of the diegetic scene and the position and orientation of the diegetic viewpoint. Chapter 3 analyzes the CHSP's from the aspects of format, image, and interaction, and infers the dynamic aesthetic experience of touring in the diegetic scenes of the CHSP.

The second part is the **Geometrization** Chapter 4 systematically analyzes the visual structure of the CHSP in different historical periods, and summarizes the formations of three image layers of the CHSP. Chapter 5 establishes the geometric correlation between the visual structure of the cell-scene image and the position and orientation of the diegetic viewpoint, and analyzes the movement of the diegetic viewpoint according to the composition of the cell-scene images in the column-scene image and row-scene image. Chapter 6 establishes the geometric correlation between the visual structure of the cell-scene image and the 3D position of the object in the diegetic scene, and analyzes the depth change in the diegetic scene according to the composition of cell-scene images in the column-scene image and row-scene image.

The third part is the **Computation & Construction**. Chapter 7 designs the algorithms for calculating the diegetic viewpoint and constructing the diegetic scene.

The fourth part is the **Synchronization** Chapter 8 builds up the viewing scene and implements the two synchronization programs in a VR application.

Chapter 9 is a discussion. It talks about the evaluations of the algorithms for calculating the diegetic viewpoint and constructing the diegetic scene, the contribution and limitation of the whole research, and future works.

Part I

Spatialization

Chapter 2

Being in a painting

This chapter will present the study on the diegetic experience of "being in a painting". First, the process of aesthetic experience in classical Chinese aesthetics will be introduced. Second, the generation process of the diegetic experience of "being in a painting" will be elaborated. Which, two transformations in this process will be explained in detail: the transformation from the 2D painting to the 3D diegetic scene, and the transformation from "watching" the painting to "being in" the diegetic scene. Third, the classical Chinese concept of worldview and the method of projection in ancient painting theory will be introduced. Based on this, this chapter will also explain the inferences of the spatial form of the diegetic scene and the viewer's diegetic viewpoint based on the combination of the above studies.

2.1 Process of aesthetic experience

Before discussing the aesthetic experience of an artwork, we shall answer a question: what exactly do we experience when we appreciate an artwork? For this question, Hegel gave a short discussion in his book of *Aesthetics*.

> "In a work of art, we begin with what is immediately presented to us and only then ask about its meaning or content. The former, the external appearance, has no immediate value for us; we assume behind it something inward, a meaning whereby the external appearance is endowed with the spirit. It is to this, its soul, that the external points. For an appearance that means something does not present itself to our minds, or what it is as external, but something else." (Hegel, 1975)

It can be concluded from Hegel's description that what we usually consider material entities, such as sculptures, paintings, etc., are not the final objects of perception in the artistic aesthetic process. The object of perception changes from the external appearance of the artwork to its internal meaning. Since the internal meaning of an artwork does not show itself directly but must arise on the basis of the perception of its external appearance, aesthetic experience is a progressive process.

In his essay of *Mingxiang* 明象, Wang Bi王弼(226 - 249) describes the process of appreciating a poem as similar to the one described by Hegel.

> The imagery represents the meaning, and the word describes the imagery. There is no more detailed representation of the meaning than the imagery and no more clear description of the imagery than the word. The imagery is generated from the

> word, so it is possible to perceive the imagery through the words. The imagery is generated for the meaning, so it is possible to comprehend the meaning through the imagery. The meaning can be expressed completely because of the imagery, and the imagery can be represented clearly because of the words. Therefore, the purpose of the word is to describe the imagery. As the one gets the imagery, he will forget the word. The purpose of the imagery is to preserve the meaning, and when one gets the meaning, he will forget the imagery. [1] (Tang, 2009)

According to Wang's discourse, the aesthetic object changes twice when appreciating a poem: from *Yan* 言(word) to *Xiang* 象(imagery) and then to *Yi* 意(meaning). The word is the initial manifestation of the meaning, and the meaning is the word's ultimate goal. Therefore, this relationship between the word and the meaning is similar to the relationship between the external appearance and the internal spirit described by Hegel. And only based on the previous level can the next level start to appear. Once the more profound level is revealed, the previous level will fade away. Therefore, the appreciation of a poem is also a progressive process.

However, unlike the direct transformation from external appearance to internal spirit, the transformation from the word to the meaning requires imagery as the medium. This difference arises not because of the different objects of interpretation, as Hegel talks about works of art in general while Wang talks about poetry, but due to the different divisions of entities in their philosophies. Hegel adheres to the dichotomy of Western Classical metaphysics. In this philosophical system, entities are generally grouped into two categories: material entities that are independent of the mind, or imaginary objects that exist within the mind. (Thomasson, 2004) For the artwork, its ontological status can only switch from the external to the internal. The aesthetic process is then a two-phrased progression.

While in traditional Chinese metaphysics, the division of entities is trichotomy. Thus, there is an additional intermediate element in the Chinese aesthetic experience. Pang (2003) summarized this division as a pattern of *Qi* 器(object) - *Xiang* 象(imagery) - *Dao* 道(principle) [2]. *Qi* is a specific thing, that which has form, function, or capability.[3] [4] It is something perceptible or something that can be described in concrete terms. However, for a thing, not everything of it can be perceived or described. Therefore, *Qi* is partial and limited. *Dao* is the metaphysical thing, that which is beyond the form.[5] It is referred to the general laws followed or the universal patterns followed by all things and beings, or the source or ontological existence of things. When we want to use a partial and limited thing to express the comprehensive and unlimited philosophy, we need to refine and extend *Qi*. This process will then lead to the generation of *Xiang*. *I Ching* 易经points out that, Xiang is the manifest form of the *Qi*。

> "The sages saw the complexity of everything in the world, so they proposed to measure the form of everything, and summarized into eight basic trigrams, to symbolize the appropriate physical image of everything, so-called 'Xiang'."[6] (Zhou, 2018)

> "In the ancient times, *Bao Xi* 包牺, who ruled the world, looked up to observe the phenomena of the sky, looked down to observe the phenomena of the earth, and

[1]Original text: 夫象者，出意者也。言者，明象者也。尽意莫若象，尽象莫若言。言生于象，故可寻言以观象。象生于意，故可寻象以观意。意以象尽，象以言著。故言者所以明象，得象而忘言。象者所以存意，得意而忘象。

[2]Dao is the Tao in Pinyin spelling. In this thesis, all translations of Chinese words will follow the Pinyin spelling rule.

[3]形而下者，谓之器。

[4]Explanations of *Qi*, *Xiang* and *Dao* 道were quoted from Terminology Library of Key Concepts in Chinese Thought and Culture. https://shuyuku.chinesethought.cn/

[5]形而上者，谓之道。

[6]Original text: 系辞上：圣人有以见天下之赜，而拟诸其形容，象其物宜，是故谓之象。

Figure 2.2.1: Xiao and Xiang Rivers. Dong Yuan. The Palace Museum, Beijing

> slowly observed the patterns of birds and animals to fit the geographical position of grass, wood, gold, and stone; taking the image of the human body in the near future, and imitating the image of all things in the far future, so he began to make the eight trigrams to integrate the wisdom and virtue of the gods and to classify all things for comparison."[7]

The Xiang is created by the sage, based on his observation of natural phenomena. But it is not self-expression of the sage, but a representation of all things in the universe. This representation is not only limited to simulating the appearance of external objects but also focuses on representing the inner nature of all things. Therefore, compared to the *Qi*, Xiang is more complete. It contains the nature of the thing represented by the *Qi*, as well as the nature that is not represented. The progressive process of the artwork in classical Chinese aesthetics then has three phases.

2.2 Being in the diegetic world of a painting

The division of a Chinese painting should follow the principle of trichotomy. The *Qi* level of painting refers to its sensible elements, including the visible forms, lines, colors, and motifs, as well as the materials, like paper and ink. The *Dao* level of the painting is reflected in the subject matter or the spiritual and emotional product of the viewer's perception of the painting. Such as the calm and relaxed feeling we would gain, when watching Dong Yuan's 董源(c.934 - c.962) painting of *Xiao and Xiang Rivers* 潇湘图. [Figure 2.2.1] The *Xiang* level of the painting generally refers to the thing or scene represented in the painting. However, due to the infinite nature of *Xiang*, the painting's *Xiang* also includes any possible imagination or extension of this thing or scene. This is like when we look at Li Cheng's 李成(919 - 967) painting of *Luxuriant Forest among Distant Peaks* 茂林远岫图[Figure 2.2.2], we not only see a three-dimensional scene which represented in the painting but also imagine the hidden scene in the mist.

The essence of the diegetic scene of the painting is its *Xiang*, and the diegetic movement in the painting is actually interacting with the painting's *Xiang*. Therefore, the implementation of "touring in the painting" is related to the generation and interaction of *Xiang*.

[7]Original text: 系辞下: 古者包牺氏之王天下也，仰则观象于天，俯则观法于地，观鸟兽之文，与地之宜，近取诸身，远取诸物，于是始作八卦，以通神明之德，以类万物之情。

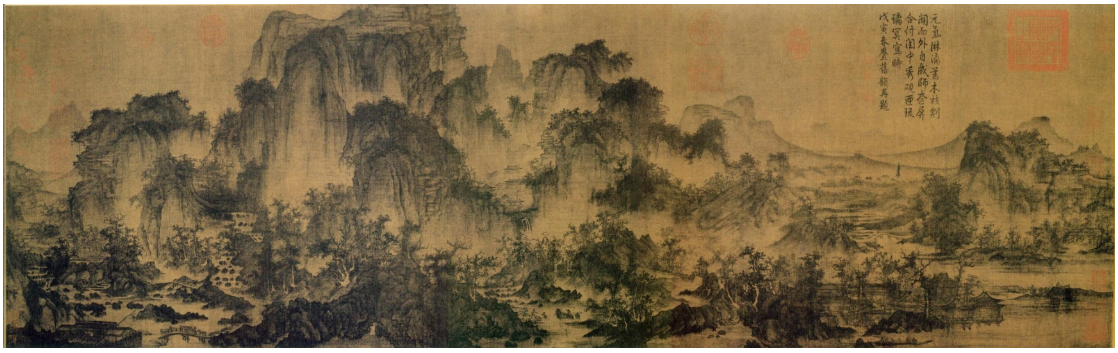

Figure 2.2.2: Luxuriant Forest among Distant Peaks. Li Cheng. Liaoning Provincial Museum, Shenyang

2.2.1 Generation of *Xiang*

In classical Chinese aesthetics, the generation of the *Xiang* is the result of an intentional act of *Guan*观. Ancient Chinese aestheticians summarized this process as *Guan Wu Qu Xiang* 观物取象(observing the object to get the imagery), which is summed up from a discourse from the book of *I Ching* 易经. The initial meaning of *Guan* is to observe and scrutinize something. From the 4th century, "Guan" became a common practice among the practitioners of the Taoist Shangqing school, and then had implications for religious cultivation. (Clunas, 1997) This implication is embodied in two characters.

First, *Guan* is meditative watching. In Taosit's theory, *Guan* is an "active, conscious reflection on one body and mind", while they are reciting scriptures. (Robinet, 1989) The practitioner will empty his mind and cast aside all kinds of worldly interference, then focus on feeling the operation of the cosmic forces in his body and the universe. (Clunas, 1997). In this way, the practitioner can host deities in all the detail of their complex and numinous iconography and conceives of religious contents, such as the appearance and activity of God, and non-religious contents, such as sky, air, etc., to form a vivid religious scene described in the scriptures. In this sense, *Guan* is a creative activity to visualize or concretize something.

Second, *Guan* is a comprehensive observation. This feature is related to the rite that the Buddhist monks practice in the caves that are decorated with murals. These murals are representations of Buddhist scenes mentioned in sutras. For example, the murals in the Cave 172 of Dunhuang Grottos represent the scenes described in the *Amitayurdhyana Sutra*. [Figure 2.2.3]Viewing such murals requires the assistance of an installation consisting of 10 or more mirrors. These mirrors are arranged around a central platform, with eight of them occupying the compass points, one above and one below, to face each other (Wang, 2004). While practicing, monks sit in the cave's center and observe the murals through the mirror installation. Since every two mirrors face each other, the murals will be reflected back and forth. In this way, images of Buddhist scenes initially scattered throughout the cave will be assembled and fused into an immersive vision of a Pure Land. In this sense, the *Guan* in the Buddhist practice became a performative rite. It integrates the scattered scenes to form an entire world of the Pure Land. The *Guan* is therefore an activity to implement something.

The influence of Buddhism and Taoism on classical Chinese painting was profound. As religious painting became widespread among the literati during the Six Dynasties (222 - 589), *Guan* thus became a meditative watching painting (Clunas, 1997). The visualization and integration features of *Guan* in religious practice were inherited and developed into two-staged intentional acts in the process of watching a painting.

Figure 2.2.3: Murals in the Cave 172 of Dunhuang Grottos. Digital Dunhuang, Dunhuang

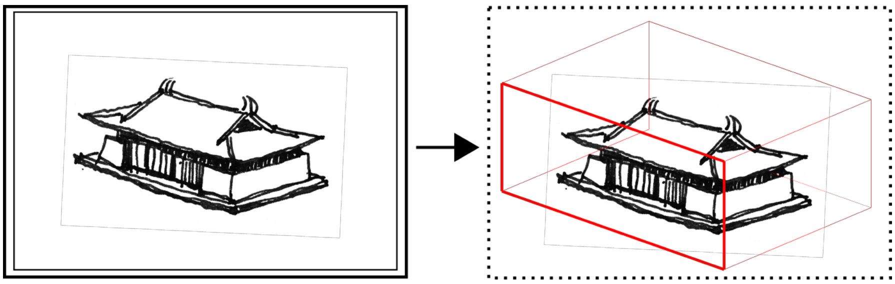

Figure 2.2.4: From 2D painting to 3D represented scene

2.2.1.1 From 2D painting to 3D represented scene

The first staged act of *Guan* is to visualize the represented object. This act is driven by our visual association mechanism. When we see the image of an object, our visual system undertakes a two steps process. First, the visual system perceives the object image, such as the forms and colors, and transmits these signals to the brain. Second, our brains engage in an imagining process to associate any experience with these signals. Because all that is experienced belongs to a unity of the self, and they all have an irreplaceable connection to the whole (Gadamer, 2013), and then these experiences are combined to gradually refine the whole. Just as when we see an image of a house, we would naturally associate it with its 3D shape, material, or size, in reality. [Figure 2.2.4]At this moment, the un-present features of the object will be integrated with those already presented in the painting as a whole (Zhang, 2013). As for a realistic painting like [Figure 2.2.4], this whole is the 3D imagery of the house.

This act is easy to understand if the represented object comes from reality. However, what if the represented object does not exist, or the viewer has not seen it at all? Then, how is this represented object visualized? In fact, the painting is the product of a high degree of concentration and emotional integration. Therefore, the experience associated with it may not come from a single thing, but rather a collection of experiences of many other things. That is, even if it is a non-realistic painting, there is always something in our experience to match

Figure 2.2.5: The mural of Kalavindotabvka in the Cave 25 of Dunhuang Grottos. Digital Dunhuang, Dunhuang

it. An example is the fresco *Kalavindotabvka* from Cave 25 of the Dunhuang Caves. [Figure 2.2.5]Creatures with the upper body of a man and the legs of a bird do not exist in reality. However, individual bird and human shapes can recall the relevant experiences of a 3D human and a 3D bird. These experiences help to form a 3D imagery of a bird-man.

Therefore, we have a great deal of experience in being able to concretize the image of an object or scene from a representational painting[8]. In this process, the object of aesthetic experience transforms from a 2D image to a 3D imagery of the represented object or a represented scene if the painting is themed on space. Classical Chinese aestheticians referred to them as *Yi Xiang* 意象(aesthetic imageries). It is to be noted that aesthetic imagery is a crystallized product of the sensation and the artwork (Zong, 2000). Its generation requires the observer's participation and it always presents in mind. Therefore, the aesthetic experience evoked by the painting may vary from person to person. However, it is not a wander less fantasy. The aesthetic imagery is implemented based on the experience of the material level of the painting. Therefore, it remains consistent with the painting in terms of the appearance that is already presented.

2.2.1.2 From partially represented scene to panoramic diegetic space

The second staged act of *Guan* is to expand the represented scene. This act is also driven by our visual association mechanism. In fact, the brain's association with related experiences is aimless. That is to say, it may recall relevant experiences about the nature of the presented object itself, such as the 3D form of the house in [Figure 2.2.4]. It may also associate with other factors related to the object but not its own nature, such as the external environment in which the house is located, or the story that takes place in the house. [Figure 2.2.6] More importantly, the brain would associate with these experiences all the time, which leads to the imagination of the represented object expanding infinitely outward in a lively state. Peng (2014) refers to Seel's "appearing" (Erscheinen) to describe this vibrant state of expansion. As something related to the represented object is infinite, the imagery of the represented object keeps expanding until

[8]Representational painting is a painting that aims to depict something real or non-real. It is distinguished from expressional painting, such as Minimalist paintings.

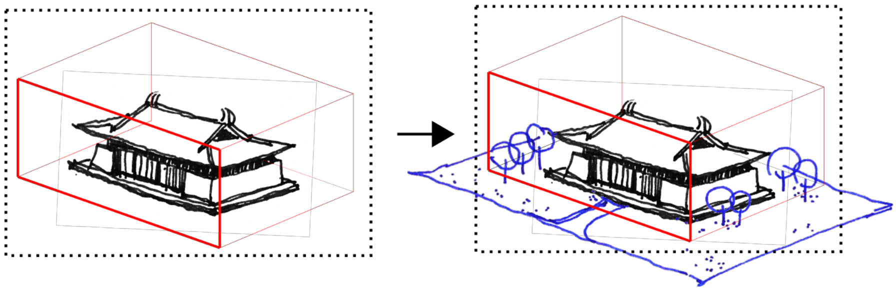

Figure 2.2.6: From represented scene to diegetic world

the entire world that is located.

However, representational painting is purposive. It wants the viewer to associate in a predetermined direction to complete the representation of a scene, such as a specific scene described in the Buddhist scriptures, rather than aimlessly fantasizing. This means that the painter needs to set up certain guiding measures to make the viewer associate along the direction he wants. In the Buddhist mural, this task was achieved with the assistance of a mirror installation. However, this method may not suitable for regular painting which represents a single image. Therefore, the ancient Chinese painters developed another method to achieve this through composition and visual guidance in landscape painting. Ma Yuan's 马远painting of *Looking at geese from a river pavilion* 江亭望雁图is a typical example of applying this approach.[Figure 2.2.7] A prominent feature of this painting is the unbalanced composition, in which the artist placed the buildings and plants in the lower right corner of the canvas while leaving a large blank space. This unbalanced composition creates a strong sense of visual dynamics. When watching this painting, the viewer will be attracted by the graphic with solid contrast and clear contour. Therefore, he will first place his focus on the well-defined area in the lower right corner. In this area, the human figures orientate to the left, which gives a hit to the viewer that there is something interesting on the left. Therefore, the viewer's focus point will secondly move to the left. In this process, the titled boat towards the upright would guide the viewer's focus point all the way to the foot of the mountain. The three mountain images form a tilted and successively diminishing sequence, which also produces a tendency of recession into the distance. Therefore, the viewer's focus will not linger here but will follow the direction of the sequence and toward the mist in the center of the painting.

Visual perception studies showed (Wang, 2015) that large blank and blurred areas in Chinese paintings could trigger greater relaxation and entering a meditative state. In this state, the viewer's mind is disengaged from the current image and begins to imagine the scene outside the image. As mentioned before, this imagination may be aimless. However, the "after-effect" (Wang, 2015) of visual perception would accumulate the visual experience acquired while viewing the painting and bring it into meditation. These experiences will act as the basis of the future imagination, guiding the viewer to imagine the larger world outside the painting based on the currently represented scene. And this process will undertake constantly, which will lead to the formation of a world in which the represented scene is located.

To distinguish this vast world from the limited represented scene, the poetry theorist Liu Yuxi 刘禹锡borrowed the Buddhist term of *Jing* 境(realm) to denote this infinite imagery as *Yi Jing* 意境(aesthetic realm). He put forward the proposition that "the realm is born outside the imagery" 境生于象外(Ye, 2009). The aesthetic realm is then a combination of the aesthetic imagery and the nihility outside of it. It is a breakthrough from finite imagery in both time and space. However, ontologically speaking, the aesthetic realm is still a *Xiang* (imagery), infinite

Figure 2.2.7: Looking at geese from a river pavilion. Taipei Palace Museum, Taipei

imagery of the diegetic world where the aesthetic imagery is located.

2.2.2 Interaction with *Xiang*

In his essay of *Preface of Painting Landscapes* 画山水序, Zong Bing 宗炳used *Cheng Huai Wei Xiang* 澄怀味象(clarify the mind to savor the imagery) to describe the interaction of *Xiang*. *Wei* 味is the intentional act with *Xiang* 象. *Cheng Huai* 澄怀is a concept related to classical Chinese philosophy[9], which refers to emptying one's mind, casting aside all kinds of worldly interference and eliminating the desire for fame and fortune. Therefore, to analyze the specific interaction with *Xiang*, we shall introduce the theory of interaction in classical Chinese philosophy.

2.2.2.1 *Qi* as the premise of interaction

The classical Chinese philosophy is Monism. In this ideological system, all things are made of the same element,*Qi* 气[10]. The literal meaning of *Qi* is air which referred to all gaseous substances. From the Spring and Autumn period (c.770 - 476 BCE) onwards, *Qi* became a philosophical concept that can be seen as the equivalent of the ancient philosophical concept of Arche (Maruyama, 1993). Lao Zi 老子made a clear explanation of the primitivity of the *Qi* in his cosmogony.

> "The *Dao* gives rise to a chaotic *Qi*, which divides into two *Qis*, *yin* and *yang*, and the two *Qis* meet each other to form a new state of harmony. All things are created out of the harmony between *yin* and *yang*. Therefore, the essence and life of all things is *Qi*, or the *Dao*. All things contain the opposing aspects or tendencies of *yin* and *yang*, but they are united in the invisible *Qi*."[11] (Liu, 1993)

[9] *Cheng Huai* 澄怀is the aesthetic interpretation of Lao Zi's "涤除玄鉴" (Wash away the dirt and dust, eliminate distractions, meditate and look deeply).

[10] This *Qi* 气has the same spelling with *Qi* 器(object), but they are different words in Chinese.

[11] Original text: 道生一，一生二，二生三，三生万物。万物负阴而抱阳，冲气以为和。《道德经》

In Lao Zi's words, *Qi* is the origin of the universe and the driving force of the evolution of all things.

Qi is also a fundamental element of the human being. As it is written in *I Ching* 易经, "*Qi* is the root of man."[12] The Taoist philosopher Zhuang Zi 庄子argued in his essay *Zhi Bei You* 知北游that "the birth of a human being is the gathering of *Qi*, the gathering of *Qi* forms life, and the dispersion of *Qi* is death."[13] Therefore, *Qi* makes human life. This idea was also shared by Confucians. In Mencius's thoughts, the human being is a trinity of *mind* 心- *Qi* - *body* 身. He believed that one's mind directly leads to his *Qi* and the *Qi* props up his body.[14] (Yang, 1996) Therefore, the human being and the universe were regarded as containers filled with the same material, *Qi*. It is the *Qi* that provides the possibility for the human being to communicate with all other things. As Wang Yangming 王阳明(1472-1529) straightforwardly pointed out,

> "Human beings and all other things between heaven and earth are original as a unity. ... It is only because of the shared element of Qi that they can communicate to each other."[15]

2.2.2.2 Resonance of *Qis* as the essence of interaction

Qi determines the way of interaction with *Xiang*. In his book of *The Record of the Classification of Old Painters* 古画品录, Xie He 谢赫of Six Dynasties proposed the concept of *Qi Yun Sheng Dong* 气韵生动(resonance of *Qis* and emotional fluctuation) to describe the reaction of *Qi* while a viewer is watching an excellent painting. The *Qis* here are respectively that of the viewer and that of the painting. Xie's concept implies that the artists of the Six Dynasties period had developed an aesthetic consensus that watching a painting is essentially an interaction between the viewer's *Qi* and the painting's *Qi*. The *Yun* 韵(resonance) is the interaction way of *Qis*. According to aesthetician Liu Xie's 刘勰explanation in his book of *Wen Xin Diao Long* 文心雕龙, the *Yuan* 韵refers to the echo of the same frequency and rhythm.[16] (Ye, 1985). As *Qi* refers to a gaseous substance, the result of the resonance of *Qis* is to make their movements with the same frequency and rhythm. Mencius believes that the movement of *Qi* would influence mental state.

> "concentration of the mind could cause changes in the movement of *Qi*, and concentration of *Qi* could trigger changes in the mind. For example, a sudden fall or a brisk walk will inevitably cause a change in the movement of *Qi*, and the original state of mind will be broken as well. Conversely, a sudden change in the mind, such as fear or grief, is also bound to cause a change in the movement of *Qi*."[17]

According to this theory, when the viewer watches the painting, his *Qi* will change because of the resonance with the *Qi* of the painting. This change in the movement of his *Qi* would further make the change in his emotion accordingly. That is said *emotional fluctuation* 生动in Xie's concept.

Xie He's 谢赫concept explains the emotional changes brought about by the interaction of the *Qi* from the viewer's side. However, the changes brought about by resonance are bilateral. Liu Xie's 刘勰*Theory of Perceive and Arousal* 感兴论gives a more comprehensive explanation of this process in the chapter on Interpretation and Poetry 诠赋of *Wen Xin Diao Long* 文心雕龙.

[12]Original text: 气者, 人之根本也。
[13]Original text: 人之生, 气之聚也。聚则为生, 散则为死。
[14]Original text: 夫志，气之帅也；气，体之充也。《孟子-公孙丑》
[15]Original text: 盖天地万物与人原是一体，......，只为因此一气，故能相通耳。《传习录》
[16]Original text: 同声相应谓之韵。
[17]Original text: 志一则动气，气一则动志，今夫蹶者，是气也，而反动其心。《孟子》

This theory consists of two core ideas: *Qing Yi Wu Xing* 情以物兴(the emotion arises from the thing) and *Wu Yi Qing Guan* 物以情观(the thing is viewed to have the feeling) (Tong, 2011). The former refers to the resonance result from the viewer's side, which means that the viewer's feeling is evoked by the thing. The latter refers to the resonance result from the painting's side. According to literary critic Zhong Rong 钟嵘(c.468 - 518), the origin of vitality to *Qi*'s movement in his monograph of *Shi Pin* 诗品, "*Qi* gives vitality to the thing, and the vitality of the thing moves the people. People's mood goes up and down, which causes dancing and singing."[18] This also implies that the thing has the ability to bear emotion. Then, if the thing's *Qi* performs the same movement as the man's *Qi* after resonance, it is supposed to have the same emotion as the man. In this sense, the resonance of the viewer's and painting's *Qis* will finally create an emotional empathy between them.

2.2.2.3 Unity of subject and object as the result of interaction

In his book of *Mikrokosmus*, Lotze believed that empathy would result in the displacement of the man's life to the thing and the displacement of the thing's life to the man (Lotze, 1897). When empathy occurs, the man sees the original inanimate thing as if it is lived, and puts himself in the situation of the thing to feel, think and act. At the same time, the man feels the effects of this illusion on the thing, empathizing and resonating with it (Zhu, 2011). Finally, empathy ultimately leads to the mutual exchange of life experiences between the man and the thing.

Such mutual exchange of life experiences was also vividly manifested in Zhuang Zi's 庄子story of *The Joy of Fish* 鱼之乐.

> Zhuang Zi 庄子and Hui Zi 惠子were walking along the dike above the River Hao when Zhuang Zi said, "Note how the minnows dart out to wander where it suits them. This is a joy for a fish." "You're not a fish," Hui Zi said, "so how do you know what's a joy for one?" "You're not me," Zhuang Zi replied, "so how do you know I don't know what's a joy for a fish?" "I'm not you, so I most assuredly don't know what you know. You're not a fish, so my proposition concerning you not knowing what is a joy for a fish remains unrefuted." "Oh, please. Let's get back to the root of the issue, "Zhuang Zi said." 'How do you know what's a joy for a fish?' That's what you asked. So you already knew that I knew when you asked me. I know it by walking above the river." [19]

The debate between Zhuang Zi and Hui Zi is essentially a debate between two paradigms of perception. The view held by Hui Zi is a dichotomy between subject and object. In his ideology, perception is the subject's apprehension and experience of the object. There is an unbridgeable divide between the subject and the object. Whereas Zhuang Zi's view is a monistic theory of *Qi*. In this philosophical framework, perception is the empathy between the subject and the object through the resonance of their *Qis*. According to Zhuang Zi's description, it is obvious that this empathy enables the subject to break through the barriers of his corporeal body to inhabit the object, and then experience the object's environment through its body. In this sense, empathy unifies the subject and the object as one. There was no subject-object division anymore. Zhuang Zi extended this unity to all things in the universe. In his treatise of *Qi Wu Lun* 齐物论, he proposed that "heaven and earth are born with me, and all things are one with me."[20] That is to say, man and all other things have the possibility of being united as one.

[18]Original text: 气之动物，物之感人，顾摇荡性情，形诸舞咏。
[19]Original text from the book of *Zhuang Zi*
[20]Original text: 天地与我并生，而万物与我为一。

This unity is purposive. Zhuang Zi said that "rest on the thing to liberate mind" 乘物以游心. That is to say, the unity of the thing and me can help to achieve spiritual freedom. He called this freedom as *Carefree* 逍遥, which is a state where the human being is freed from the constraints of his physical body and embedded in the thing and then explores freely in the realm of infinity by following the laws of all things and harnesses the changes of the six *Qi*.[21] The experience of *Carefree* was seen as the highest level of its aesthetic experience of artworks (Ye, 2014).[22]

From now on, a conclusion on the interaction of the painting's *Xiang* can be made based on the analyses above. According to the analysis in the previous stage, the *Xiang* of the painting is the diegetic world in which the represented scene is located. Then, the interaction with the *Xiang* is the interaction between the viewer and the diegetic world. According to the essence of the interaction of subject and object in classical Chinese philosophy, the perception of the diegetic world is the resonance between the *Qis* of the viewer and the diegetic world. This resonance will keep the movement of their *Qis* the same rhythm, which will further lead to empathy between the viewer and the diegetic world. According to Zhuang Zi's theory of *Carefree* 逍遥, this empathy means that the viewer and the diegetic world have been unified as one. The viewer then can mentally liberate himself from the constraints of the physical body, breaks through the boundary between the diegetic world and the real world, and finally enters the diegetic world. That is to say, the interaction with the painting's *Xiang* finally achieves a paradigm shift from visually watching the painting outside to bodily experiencing its diegetic world inside. Just as Dong Qichang's 董其昌(1555 - 1636) commented on Fan Kuan's 范宽(c.960 – c.1030) paintings, [Figure 2.2.8]

> "sitting and watching a painting, clouds, and smoke suddenly arise ...... every time I watch them, I lose myself in thousands of rocks and ravines."[23]

2.3 Spatial shape of diegetic world & position and orientation of diegetic viewpoint

After the analysis of 2.1 and 2.2, we have derived the generation process of the diegetic world and the access process of the viewer from the real world to the diegetic world. But we have not yet known the spatial form of this diegetic world and the position and orientation of the viewer in this world. These two issues are related to two concepts in ancient Chinese culture: the concept of the world and the concept of representation.

According to Tuan Yifu (1977), the ancient Chinese worldview is highly anthropocentric and empirical. He summarizes the worldview of the ancient Chinese as the diagram shown in [Figure 2.3.1-Left] by referring to the pattern of fourfold gods on the roof brick of the Han Dynasty. In this pattern, the man is placed at the center. Four edges of the pattern stand for four basic directions and seasons, each of which corresponds to a color and an animal.

> Close to the east edge is the Blue Dragon, which stands for the color of vegetation and the element wood. Occupying the direction of the rising sun, it is also a symbol of spring. To the south is the Red Phoenix of summer and of fire with the sun at its

[21] Original text is from Zhuang Zi's article of carefree traveling 逍遥游. "If one follows the nature of all things in the universe, harnesses the changes of the six *Qis*, and travels in the infinite realm, what else does he need to rely on?" 若夫乘天地之正，而御六气之辩，以游无穷者，彼且恶乎待哉！

[22] Zhuang Zi's philosophy was highly esteemed in classical Chinese aesthetics. The unity of the thing and me and the unity of the mind and the body. 身心合一are considered as two ultimate ideals of classical Chinese aesthetics.

[23] Original text: 凝坐观之, 云烟忽生. . . . . . 每对之, 不知身在千岩万壑中。

Figure 2.2.8: Traveller among mountains and streams. Fan Kuan. Taipei Palace Museum, Taipei

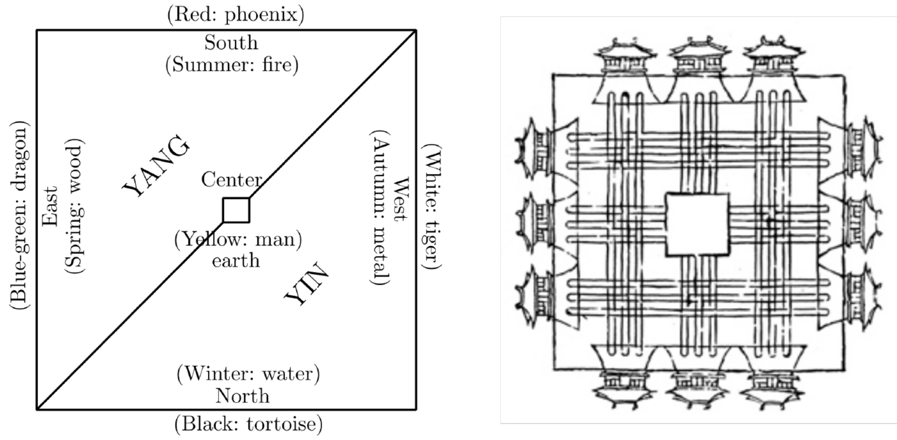

Figure 2.3.1: Ancient Chinese worldview; The Map of King's City

> zenith. To the west is "the White Tiger of the metallic autumn, symbolic of weapons, war, executions, and harvest; of fruitful conclusion and the calmness of twilight, of memory and regret, and unalterable past mistakes." To the north is winter's darkness, out of which all new beginnings must come. North is associated with reptiles, black color, and water. At the center of the cosmos is the man on the yellow earth. (Tuan, 1977)

In addition, the left side of the diagonal line is Yang (the world of the living, the bright side, positive and happy), and the right side is Yin (the world of the dead, the dark side, negative and sad). As can be seen from this diagram, the ancient Chinese worldview is a man-centered space-time system, in which space and time extend outward from man based on their empirical experience.

It is on this worldview that the ancient Chinese constructed the world. One of typical example is the ancient ideal city illustrated and described in the book of *Kao Gong Ji*考工记. [Figure 2.3.1-right] The front view of the building on this map indicates that the designer's position is located in front of the building and looking at the building. Then, the position that can satisfy four sets of centrifugally arranged front views at the same time must be located in the center of the city and looking outward. Therefore, the designer designed the city with himself at the center. In addition, the book provides further explanations of this map.

> " An ideal city should be surrounded by nine Li (里Chinese length unit) long wall on each side, with three gates on each wall. there should be nine south-north roads and nine east-west streets inside... the palace in front, while the market at the back; ancestor' temple on the left and the state altar on the right..."[24] (He, 1996)

The orientation terms "left and right" and "front and back" in the text indicate that the designer defined the direction according to his own orientation.

Since the conception of the diegetic world of a painting is essentially a construction, it is also inevitably under the influence of this worldview. Then, the diegetic world the viewer conceived can be inferred as an infinite time-space centered on the viewer and expanding in all directions. Combined with the analysis in 2.2.1, the scene of the diegetic world contains two parts. One is

[24] Original text: 匠人营国。方九里，旁三门。国中九经九纬，经涂九轨。左祖右社，面朝后市，市朝一夫。

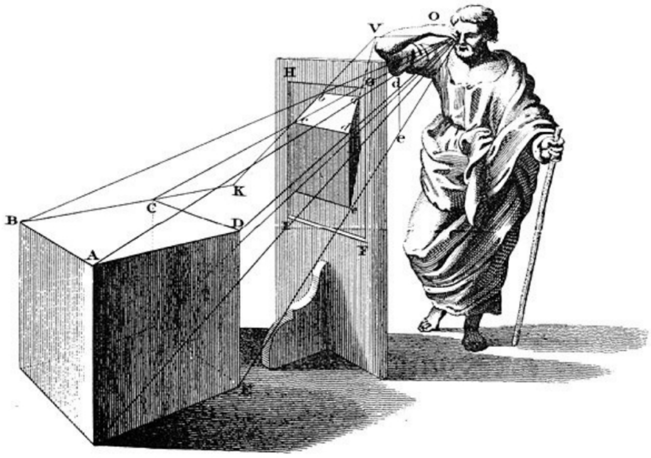

Figure 2.3.2: Setting of perspective. Taylor Brook

the scene represented in the unrolled frame. The other is the rest scenes of the diegetic world imagined by the viewer, which refers to the scenes that are not depicted in the unrolled frame.

Related to the position and direction of the viewer is the ancient Chinese concept of representation. In the article of *Preface to Painting Landscapes* 画山水序, Zong Bing 宗炳(375 - 443) gives a brief exposition of the concept of representation.

> I capture the image of the landscape through my eyes and get a sense of it in my mind, which gives me the nature (*Dao*) of the landscape. If the painting was done delicately, what the viewer sees in the painting would be the same as what the painter saw while he was in the landscape, and what they think would be the same as well. [25]

From Zong Bing's words, it is clear that painting is used as a medium to convey the painter's vision to the viewer. And according to Zong Bing's description of the viewing way that "stretching the silk to project the far scenery" in the same article, the painter looks at the scene through semi-transparent silk, which is very similar to the setting of perspective projection depicted by Brook (Brook, 1963). [Figure 2.3.2]From this, it can be inferred that the real scene seen by the painter through the silk is the same as the diegetic scene seen by the viewer through the canvas. When this situation is satisfied, the relative position and orientation of the painter's viewpoint to the real scene are consistent with the viewer's position and orientation to the diegetic scene.

Finally, we should make a summary of the spatialization of the diegetic experience.

According to the analyses above, the experience of "being in the painting" can be illustrated in the diagram in [Figure 2.3.3]. The sphere represents the whole scene of the diegetic world. The center of the sphere is the position of the viewer. The part of the scene along the viewer's sight-line is the scene (aesthetic imagery) that is represented in the painting. While the scene outside of the represented scene is the aesthetic realm. They are imagined by the viewer based on the represented scene.

[25] Original text: 夫以应目会心为理者，类之成巧，则目亦同应，心亦俱会。

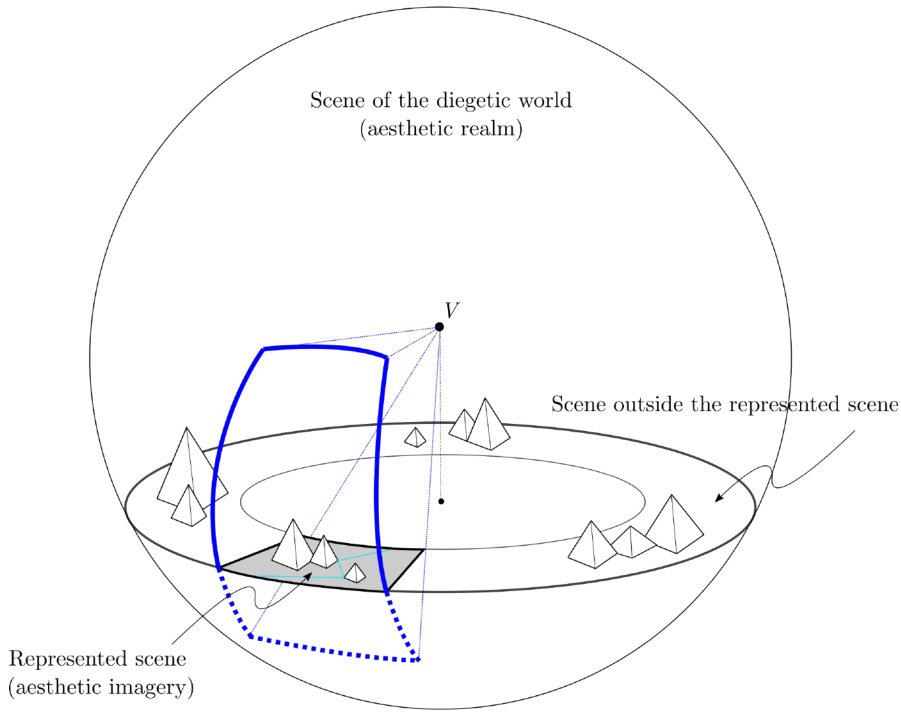

Figure 2.3.3: The scene of the diegetic world (aesthetic realm)

Summary

This chapter reviewed the aesthetic process of painting in classical Chinese aesthetics. This process is progressive and phrased. The number of phrases relates to the trichotomy division of entities in Chinese philosophy. According to this division, the aesthetic process of painting can then be divided into three phases. And the experience of being in the diegetic scene of the painting happen in the first and second phrases. The first phase is the viewer perceiving the *Qi* of the painting and then generating the *Xiang*. The intentional act of this phrase is *Guan*. *Guan* is originally a religious practice. It is a way of visualizing the scriptures and integrating the various represented scenes into a world. As religious paintings became popular among the literati, *Guan* became a meditative way to view paintings. The visualization and integration of *Guan* were inherited and developed into two-staged intentional acts. The first staged act of *Guan* is to evoke all related experiences about the 2D painting and then form the 3D imagery of the scene depicted in the painting. The second staged act of *Guan* is to integrate other experiences related to 3D imagery to expand its scene to the entire world where it is located. The second phrase is the viewer interacting with the *Xiang* of the painting. In classical Chinese aesthetics, the essence of the interaction between subject and object is the resonance of their *Qis*. Resonance makes their *Qis* keep the same frequency and rhythm of movement, which further leads to empathy between the subject and the object. The realization of empathy means that the subject and the object have been unified as one. The subject is then freed from his physical body and shares life experiences with the object. As for the painting, its imagery is the diegetic world. The interaction of the viewer and the diegetic world is the resonance of their *Qis*. The ultimate result of this resonance is the viewer being mentally liberated from the constraint of his physical body and being in the diegetic world.

According to the ancient Chinese worldview a man-oriented time-space system and the concept of representation in ancient Chinese art, the diegetic world is then can be inferred as a spherical scene centered on the viewer and expanding in all directions. This world consists of two parts. One part is the represented scene, which corresponds to the image in the unrolled frame. It appears in the direction of the viewer's sight. The other part is the rest of the scene conceived for the viewer, but not described in the active frame. It surrounds the viewer.

Chapter 3

Touring in a Chinese Hand-Scroll Painting

This chapter will present the study of the CHSP and the inference of the diegetic experience of "touring in the CHSP". First, the CHSP will be fully examined from three aspects: form, image, and interaction. The peculiarities of the image and interaction of CHSP under the influence of the form will be highlighted in this examination. Second, this chapter will explain the inference of the diegetic experience of "touring in the CHSP". In this explanation, the features of this diegetic experience because of the particularities of the CHSP will be elaborated.

3.1 Form of CHSP

The CHSP is one of the most representative types of classical Chinese painting. The origin of the CHSP can be traced back to the ancient bamboo scroll book. The painting theorist Zhang Yanyuan 张彦远(c.815 - c.877) documented a story about the birth of the CHSP in his book of *Famous Paintings through History* 历代名画记. The Emperor Mingdi 汉明帝of Eastern Han ordered his royal painter to add illustrations to the left of the text in the scroll book. This type of scroll book with illustrations was the prototype of the CHSP. However, it was not until the Northern and Southern Dynasties period that such scroll was officially used as a specialized form of painting. (Xue, 1993).

In the Chinese context, the CHSP is often called scroll painting 图卷or simply scroll 卷. However, the fact is that scroll painting is a generic term for all the paintings mounted on long sheets of paper or silk and have to be rolled up when stored. Depending on the rolling direction, scroll paintings can be classified into two categories. One is the Vertical-Scroll 立轴, which has to be rolled in the vertical direction. The other is the Horizontal-Scroll 横轴which has to be rolled horizontally. The artist Mi Fu 米芾(1051 - 1107) of the Song Dynasty further subdivided the Horizontal-Scroll into the Hanging-scroll 横挂and the Hand-Scroll 手卷, depending on their sizes and displaying ways. The Hanging-Scroll is the one less than three *Chi* 尺(94cm) long and needs to be fully open and hung on the wall. In contrast, Hand-Scroll is usually longer than three *Chi*. As its name indicates, the viewer needs to hold the Hand-Scroll in the hands and unroll it gradually while watching it.

Over the centuries, the Hand-Scroll Painting eventually evolved into the standard form in [Figure 3.1.1].Silbergeld (1982) gives it a precise description.

"Hand-scroll paintings range from less than three feet to more than thirty feet

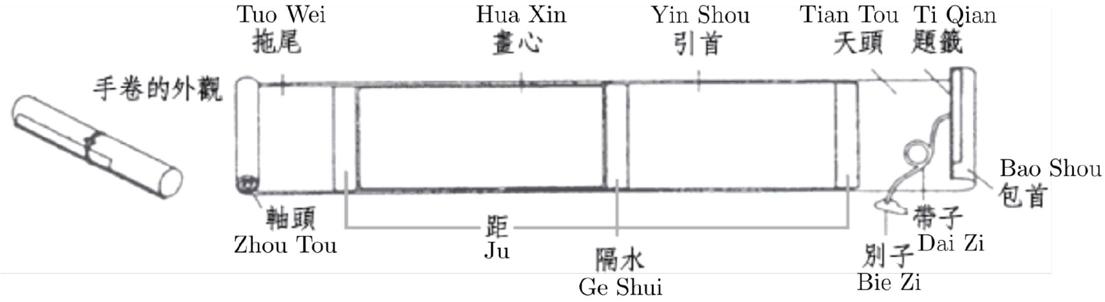

Figure 3.1.1: Form of the Chinese Hand-Scroll Painting

in length; the majority are between nine and fourteen inches high. Paintings are mounted on a stiff paper backing; those of greater length are often painted on several sections of silk or paper joined together. At the left is attached a round wooden roller, about which the scroll is wound when not in use and is occasionally decorated with a knob of ivory or jade. At the right is a semi-circular wooden strip that keeps the scroll properly stretched from top to bottom." (Silbergeld, 1982)

In summary, the CHSP is a kind of painting with a special material form. This form gives it some special capabilities in representation. First, the CHSP has a long canvas, which is suitable for the horizontal combination of multiple images. Second, the CHSP has a scrolling format that can only be unrolled a portion at a time, which makes the image show up gradually. These characteristics of the CHSP make it a considerable medium for image narrative.

3.2 Image of CHSP

The subject of this painting is related to its function in society. Zheng (2001) divides the history of ancient Chinese painting into four periods according to the painting's social functions.

1. Period of pragmatism (prehistory)[1]
2. Period of ethics and patriarch morals (from Xia Dynasty to Han Dynasty)
3. Period of religionization (from the end of the Han Dynasty to the Tang Dynasty)
4. Period of literalization (from Song Dynasty to Qing Dynasty)

The CHSP was developed and flourished in the periods of religionization and literalization, during which the CHSP undertook the tasks of enlightenment and self-expression. These social requirements determine the subjects of the CHSPs.

The themes of the CHSPs during the religionization period are moral and religious stories. A typical example is the painting of *Admonitions of the Court Instructress* 女史箴图by Gu Kaizhi 顾恺之(c.344 - 406) depicts 12 stories of virtuous concubines from history. [Figure 3.2.1] This painting was used to teach the ladies in the emperor's harem to follow the doctrine of women. Another example is the painting of *Born of Gautama Buddha* 送子天王图drawn by Wu Daozi 吴道子(c.685 - 758) based on the Sutra on the *Life of the Prince in Accordance with Good Omens.* [Figure 3.2.2] This painting represents a Buddhist scene in which Siddhartha Gautama's father, King Jokhang, and Lady Moya carry him on a pilgrimage to the temple of the Mahesvara, where the gods worship him. The ancient Chinese used it as an illustration of the Buddhist sutra.

[1]See Appendix A - Chronological table of Chinese dynasties

Figure 3.2.1: Admonitions of the Court Instructress. British Museum, London

Figure 3.2.2: Born of Gautama Buddha. Osaka City Museum of Fine Arts, Osaka

Figure 3.2.3: Clear weather in the valley. Museum of Fine Arts, Boston

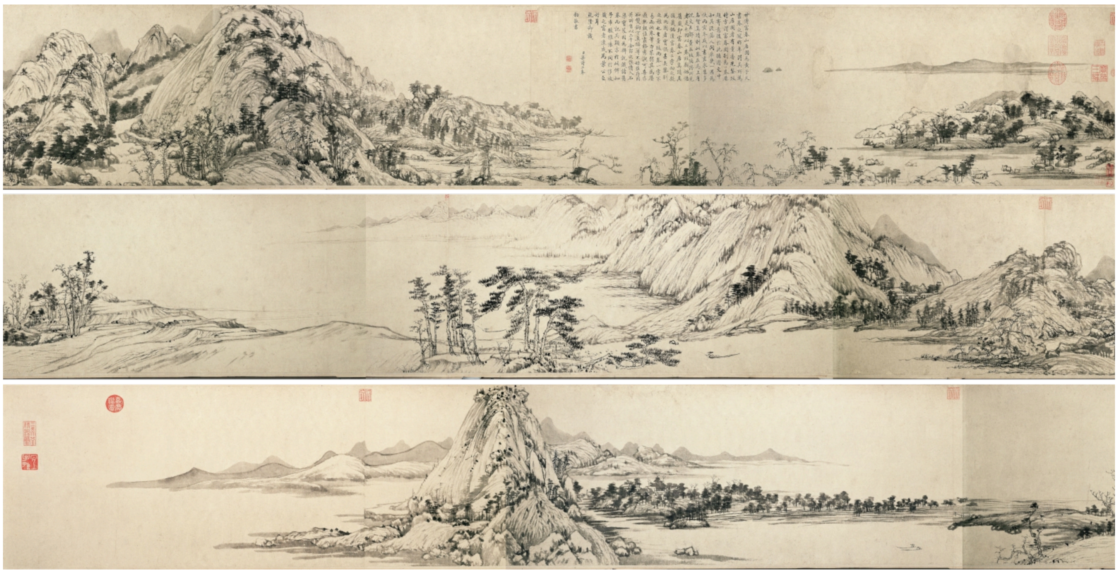

Figure 3.2.4: Dwelling in the Fuchun Mountains. Taipei Palace Museum, Taipei

From the Northern and Southern dynasties onwards, Taoist "out-of-this-world" philosophy began to be promoted by literati painters. They were no longer enthusiastic about educating people or reforming society through painting but began to express their emotions and spirituality. Consequently, the landscape began to serve as a metaphor for this task and soon became a dominant subject in the CHSPs. One of the typical works is Dong Yuan's 董源(c.934 - c.962) *Clear weather in the valley* 平林霁色图. [Figure 3.2.3] This painting represents two different scenes: a dense mountain forest and a vast water area. The mountains and forest have clear brush strokes and rich layers, presenting a deep and lush feeling. While, the ink of the water is hazy and the mountains' silhouettes are gentle, giving a spacious and calm feeling. Through the two contrasting scenes, the painter expressed his mixed mental state of tranquil, calm, and serene. Another typical work is the painting *Dwelling in the Fuchun Mountains* 富春山居图by Huang Gongwang 黄公望(1269 - 1354). [Figure 3.2.4] This painting represents the scenery changes on the banks of the Fuchun River from flourishing early autumn to withering late autumn. The painter used this change as a metaphor for the highlights and lassitude in life. The calm water at the end of the painting symbolizes his peaceful state of mind after he has lost sight of fame and fortune.

No matter telling a story or expressing a change in the mental state, it is often realized by narrating multiple scene images. In the existing CHSPs, there are two kinds of compositions. One type is compartmental juxtaposition. A typical example is the painting of *Admonitions of the Court Instructress*. In this composition, the scene images are composed independently. Clear boundaries are retained between them. The other type is open fusion. In this composition, the boundaries of the scene images are eliminated. Their individual image spaces are fused into a larger unified image space. This composition is well demonstrated in Gu Kaizhi's 顾恺之(c.344 - 406), another work of *Painting of Goddess Luo Rhapsody* 洛神赋图, which illustrates the romantic story of Cao Zhi 曹植and Godness Luo 洛神, from their encounter to their farewell. [Figure 3.2.5] The entire painting is composed of 12 scene images that are distributed throughout the canvas in chronological order from right to left. Gu removed the boundaries of the scene images and fuse them into a unified background with the continuous river banks. This manipulation creates

Figure 3.2.5: Goddess Luo Rhapsody. Gu Kaizhi. Liaoning Provincial Museum, Shenyang

the visual effect that the scenes in different time-spaces coexist in a continuous space. Because of its continuousness, this composition was largely adopted in landscape CHSPs. But unlike the *Painting of Goddess Luo Rhapsody* where each scene image has a vertical annotation next to it that can be used to divide sub-scenes, there is no obvious reference to divide sub-scenes in the landscape CHSP. Therefore, this kind of CHSP was easily recognized as a panoramic representation of a single space from a fixed viewpoint like a European panorama. [Figure 3.2.6]

However, Zhao Mengfu's 赵孟頫 (1254 - 1322) painting of *Autumn Colors on the Que and Hua Mountains* 鹊华秋色图proves that the fact is not the case. [Figure 3.2.7]This CHSP depicts the autumn scenery in the area of Mount Hua and Mount Magpie in the northeast of Jinan. [Figure 3.2.8-top]Judging from the painting, these two mountains are close to each other. However, in the panoramic photo in [Figure 3.2.8-bottom][2], it is hard to find the mountain on the left. While in fact, the missing mountain is several miles away.This phenomenon suggests that this CHSP is actually an artificial composition of scene images at different locations.

This inference is supported by ancient painting theories. In his book of *The Record of the Classification of Old Painters* 古画品录, Xie He 谢赫(around 550) proposed two principles of Chinese painting: "conform with the objects to give likeness 应物象形" and "plan and design, place and position (i.e., composition) 经营位置"[3]. This former principle means that the painter needs to depict objects according to their forms[4]. The latter refers to composing and positioning objects in the painting. The co-emergence of these two principles means that Chinese landscape painting is not a completely realistic representation of natural space. It is realistic at the level of individual objects or single scenes. But at the level of composition, the painter's subjective

[2]http://www.zeng-han.com/index.php?/new-project/-real-shanshui/

[3]Total principles of Chinese painting has six items. The rests are, Spirit Resonance (or Vibration of Vitality) and Life Movement 气韵生动; Bone-manner (i.e., structural) use of the brush 古法用笔; Apply the colors according to the characteristics 随类赋彩; To transmit models by drawing 传移模写. Translation from Siren (2005).

[4]The form here should be the objective shape of the object but not exactly its visual forms in vision. The distinction between objective and visual forms can be analogized to Gombrich's (2000) description of the difference between ancient Egyptian and ancient Greek images of ponds.

Figure 3.2.6: Robert Barker's panorama, City Art Center, Edinburg; Reconstruction of Robert Barker's panorama, Laurent Lescop

Figure 3.2.7: Autumn Colors on the Que and Hua Mountains. Zhao Mengfu. Palace Museum, Taipei

Figure 3.2.8: Photos of Mont Magpie and Mont Hua; Panorama of Mont Hua, Zeng Han

manipulation intervenes in the process for his expressive needs.

Based on the analysis above, we can make inferences about the formation process of the CHSP's images. The painter most likely captured images of different scenes from different positions, and then stitched them together in a horizontal frame, as shown in [Figure 3.2.9]This making process is very similar to the montage technique used in cinematography. Therefore, the CHSP's image can be seen as a film edited by the painter.

3.3 Interaction with CHSP

In his book of *Chinese Painting Style: Media, Methods, and Principles of Form*, Silbergeld describes the interaction of CHSP as follows.

> "The painting is viewed from right to left, as one reads in Chinese, unrolling a bit at a time from the roller and transferring the excess to a loose roll temporarily maintained around the stretcher on the right. About one arm's length is exposed at a time for viewing." (Silbergeld, 1982)

This short paragraph points out the two main dynamic acts of watching a CHSP: scrolling and "reading".

The emergence of the scrolling act is due to the long-form of CHSP, as it could not be fully opened but only unroll apart at a time. As described by Silbergeld, the basic manipulation of CHSP is a set of alternating actions. The viewer needs to unroll a part about arm's length from the side of the wooden roller and then watch this part. After watching, he needs to roll up this seen part, and then simultaneously unroll an unseen part. [Figure 3.3.1]In addition, it is necessary to add that the viewer is completely free to manipulate the CHSP. He can stop at any place and watch it carefully, or he can roll back to see what he has already seen, and watches it back and forth. (Wu, 1996) He can unroll a part of it at once and then roll it up completely, making the scene images appear and disappear as a whole. He can also scroll both the roller and strip in one direction, so that scene images appear gradually with transitions. If we consider the displayed part of the CHSP between the roller and the strip as an unrolled frame, then these actions would create a dynamic visual effect that the scene images in the CHSP are sequentially presented in the unrolled frame. [Figure 3.3.2]This visual effect is very similar to the effect of projecting different images onto the screen in a film screening. Therefore, if we follow the idea of metaphorizing the making process of CHSP as montage, the viewing process of CHSP is equal to the screening of a film.

However, the two narrative ways of the images are different. As for the film, the image narrative is controlled by the director. While for the CHSP, the image narrative is directed by the viewer. The difference in the controlling subject is decisive for the final result of the narrative. In the film, the director can control the direction and speed of the screening to recreate a preset narrative. During the whole process, the viewers passively experience the same narrative. While watching a CHSP, the viewer was active. He is free to control the speed of the scrolling, and even change the scrolling direction. This autonomy means that there is no pre-determined narrative when watching CHSP. The image narrative experienced by different viewers in different states will be unique.

The act of "Du" 读(reading) is attributed to the scroll book in which the form of CHSP originated. Silbergeld (1982) indicates that the CHSP is generally viewed from right to left, which is also the direction in which ancient Chinese people read books. In fact, according to Clunas (1997), *Du Hua* 读画(reading painting) was indeed a method of viewing paintings adopted by the literati in the Ming Dynasty.

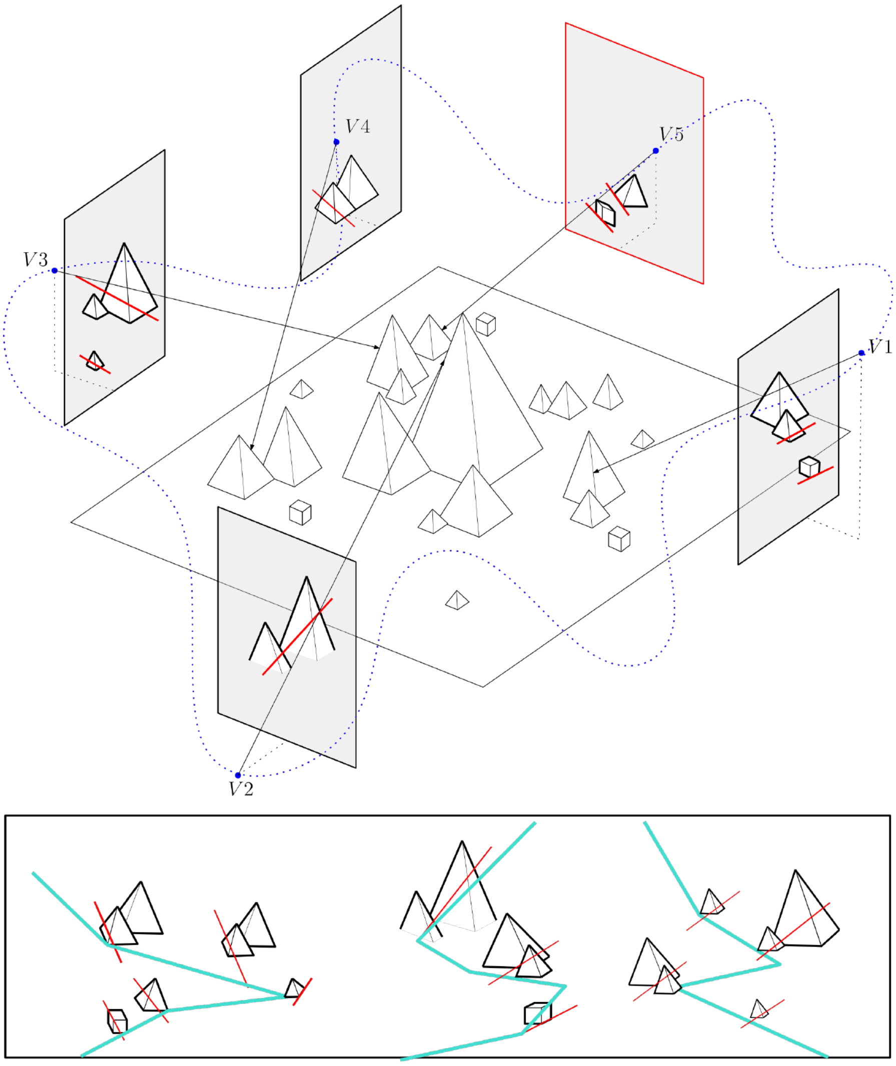

Figure 3.2.9: Formation process of the CHSP's image

Figure 3.3.1: Interaction of the Chinese Hand-Scroll Painting. The New York Times

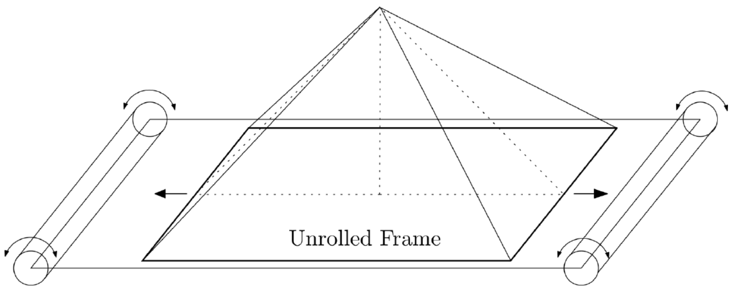

Figure 3.3.2: Unrolled frame of the CHSP

> "The idea of *du*, reading, above all implies a subject whose vision is moving, scanning the characters of a text or the surface of a picture. Importance was attached not to the legibility of the image but to the act of moving the eye across the surface, particularly of the hand scroll, as it is sequentially made visible in the act of unrolling, the presence of duration in Ming ideas of visuality is important here, the idea that pictures could not by their physical nature be taken in all at once." (Clunas, 1997)

From Clunas's description, it can be concluded that reading has the characteristics of dynamics and continuity. Since the reading needs to be done with the painting unrolled, this dynamism is reflected in the movement inside the unrolled frame.

By reading, the viewer can complete the narrative within the same image. Our eyeball is similar to a zoomable camera. When we want to look at a painting carefully, the focus of the eye would contract to a small area centered on the focus point. In this case, the part of the painting located in the region would be clear, while the rest of the painting would be relatively blurred. This region moves when the focus point moves, which would produce a visual phenomenon that the clear area alternately changes but the overall image remains. This approach is similar to that of cinematography, in which objects of different depths and positions in the same scene are alternately clear by adjusting the focus and focus position. This way of viewing is more prominent for CHSPs that have dense combinations of scene images. Such as Zhang Zheduan's 张择端(1085 – 1145) painting of *Along the River During the Qingming Festival*. [Figure 3.3.3]When the viewer unrolls an arm's length of the painting, multiple scene images will appear in the unrolled frame at the same time. Then, as the focus point moves, the scene image that appears in the focus area will gradually change, which is equivalent to switching the scene images within the unrolled frame.

3.4 Spatial narrative in the diegetic world and diegetic movement

The aesthetic experience of painting is the result of a synthesis of its form, image, and interaction. In the Chapter 2, the general aesthetic process of an ordinary classical Chinese painting has been analyzed. The ordinary paintings mentioned here are the kind of paintings with fixed frames. Its image is also fixed and can usually be viewed all at once. The viewer usually stands or sits still in front of the painting and gazes at it from a certain distance. Therefore, the aesthetic experience corresponding to such a painting is a relatively static experience of "being in the painting".

Compared with ordinary paintings, CHSP has certain special features in terms of form, image, and interaction, which inevitably bring different aesthetic experiences. The first particularity of

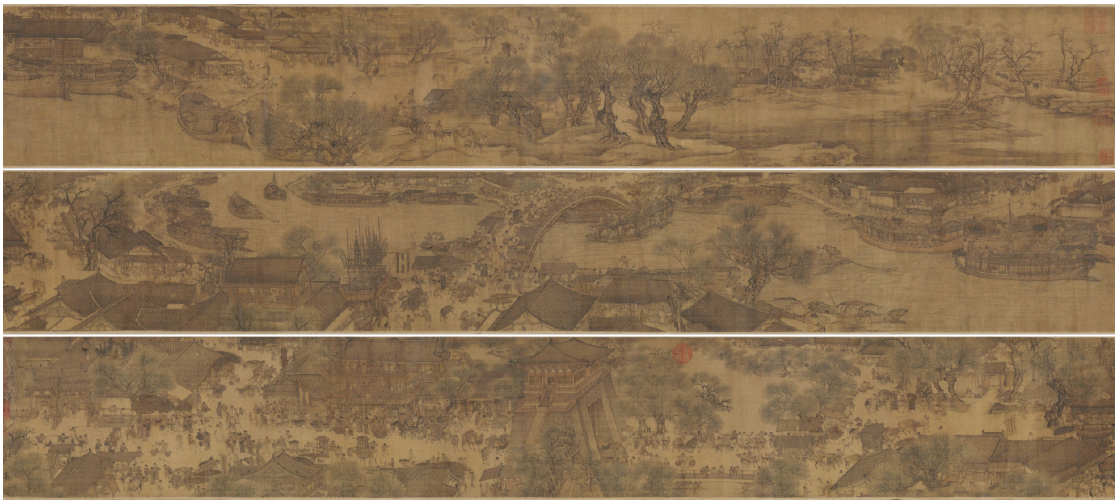

Figure 3.3.3: Along the River During the Qingming Festival. Zhang Zheduan. The Palace Museum, Beijing

the aesthetic experience embodies in the spatial narrative. According to the scrolling manipulation of the CHSP, the viewer can only unroll one arm's length of it at a time. When the viewer stops scrolling to see the displayed image, it will be perceived as an ordinary painting as it has a fixed frame. Accordingly, he will experience a complete aesthetic process of an ordinary painting. In this process, the 3D represented scene and its diegetic world will be conceived corresponding to the image in the unrolled frame. However, the images in the unrolled frame will change, as the viewer scrolls the CHSP. This change of the image in the unrolled frame will instantly trigger a new round of the aesthetic process. Then, the represented scene and the diegetic world corresponding to this image will also change accordingly. In this sense, the panoramic space experienced by the viewer is in the narrative. Also, as the change of the image in the unrolled frame is completely controlled by the viewer, this spatial narrative is also directed by the viewer. According to the analysis in 3.3, the viewer can manipulate CHSP in multiple ways. In most cases, the viewer would not strictly switch the displayed images by scene image, but would gradually switch them by continuously scrolling the roller, especially for landscape CHSP, where the scene images are continuously fused without clear boundaries in between. This causes the image in the unrolled frame to fade in and out gradually, which would create a continuity of content between the sequentially displayed images. Accordingly, the continuity between images also arises in the synchronous continuities in the represented scene and the panoramic spaces in the diegetic world. At the same time, the human visual perception has "after effects" that the previous vision will be retained for a short period a Therefore, the aesthetic experience of the CHSP is a continuous spatial narrative.

The second particularity of the aesthetic experience reflects in the diegetic movement. According to the analysis in 3.2, the image of CHSP is a combination of multiple scene images. Each of them corresponds to a diegetic viewpoint with a particular position and direction. In most of CHSPs, the lengths of these scene images are usually less than one arm's length. This would create a frequent phenomenon in which multiple scene images coexist in the unrolled frame. When the viewer adopted the approach of "Du" (reading) to watch a CHSP, his vision will continuously move across these scene images in the unrolled frame. As these scene images correspond to different locations and orientations, the aesthetic imageries will sequentially ap-

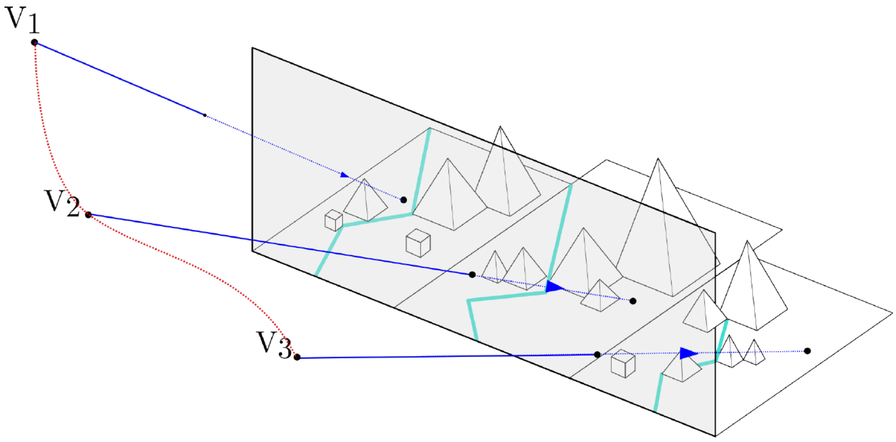

Figure 3.4.1: Spatial narrative of diegetic scenes

pear in different positions and face the viewer with different orientations. From the viewer's perspective, this experience is equivalent to moving from one position to another and watching different scenes from different directions. [Figure 3.4.1]

The above are the particularities of the CHSP aesthetic experience. Their synthesis generates of the diegetic experience of "touring in the CHSP".

Summary

In this chapter, we analyzed the CHSP from the aspects of form, image, and interaction.

The CHSP is one type of scroll painting with a long and narrow form, which originates from the ancient bamboo book. It was formally established as a painting specialized form in the Southern Dynasties and had become the most popular type of classical Chinese painting since the Jin Dynasty.

CHSPs' themes are mostly religious and historical stories during the period of religionization. From the period of literalization, the landscape became mainstream.

The image of the CHSP is essentially a composition of multiple scene images. These images are commonly composed in two ways: compartmental juxtaposition and open fusion. In the former composition, the scene images are juxtaposed with closed boundaries. The images in the latter composition are fused in a unified pictorial space. This composition was widely applied in the CHSP of landscape to create a continuous visual effect.

Watching a CHSP need two dynamic actions: scrolling and "Du" (moving vision). Scrolling transforms a still painting into a dynamic image narrative. "Du" enables switching the area of interest while keeping the image in the unrolled frame unchanged.

The aesthetic experience of CHSP is in line with the general aesthetic experience of ordinary painting but has two specificities. One is the continuous spatial narrative produced by the scrolling manner. The other is the experience of traveling across different scenes produced by the moving vision.

Part II

Geometrization

Chapter 4

Visual structure of Chinese Hand-Scroll Painting

The chapter will elaborate on the study of the visual structure of the CHSP. The elaboration will follow the three-layer division of CHSP's images in 1.3.2 and unfold layer by layer. First, the primordial form of the visual structure of CHSP will be introduced. Second, this chapter will explain the visual structure of the cell-scene image and the inference of its formation. In this explanation, ancient painting theory and architectural graphical theory will be introduced. Third, the three types of visual structures of column scene image and row scene image will be explained respectively. In this explanation, the CHSPs that have typical compositions of three development stages will be used as examples for explanations.

4.1 Parallelogram as the primordial form of visual structure

In his article of *Why Chinese painting is history*, Fong indicates that the diagonal line is a primitive finding in pictorial composition, positioning the figure as a parallelogram that may indicate a spatial recession (Fong, 2003). In the existing CHSPs, such parallelograms exist in three ways. The most obvious way is to present the parallelogram directly as a part of the visual structure of the object image. Such as in the images of furniture or a building in Qiu Ying's 仇英painting of *Spring Morning in the Han Palace* 汉宫春晓图, the visual structures of the top or bottom faces of furniture and buildings appear as parallelograms. [Figure 4.1.1]Another common way is to imply parallelograms by the structural relationship of object images. This case can be found in the painting of *Court Ladies Preparing Newly Woven Silk* 捣练图by Zhang Xuan 张萱(713-755). [Figure 4.1.2]The ladies at either end of the painting are placed around a square object, silk, or a pounding table. In each scene, the two of them in the middle create staggers in both horizontal and vertical directions. Then, the four ladies' position frame a closed parallelogram area from the painting. This method was also applied to landscape painting. Such as the painting of *Recumbent Journey in Xiaoxiang* 潇湘卧游图. [Figure 4.1.3]The triangular motifs in the painting, arranged in a staggered and overlapped manner, will be seen as a sequence with an oblique gradient. Then the hidden lines connecting the shoulders on both sides of the mountain motifs imply the existence of a parallelogram. The most implicit way to imply the parallelogram is by the front-and-back composition. This composition was adopted in the painting of *Admonitions of the Court Instructress*. In the selected image [Figure 4.1.4], two

Figure 4.1.1: Spring Morning in the Han Palace. Qiu Ying. Taipei Palace Museum, Taipei

Figure 4.1.2: Court Ladies Preparing Newly Woven Silk. Zhang Xuan. Museum of Fine Arts Boston, Boston

Figure 4.1.3: Recumbent Journey in Xiaoxiang. Tokyo National Museum. Tokyo

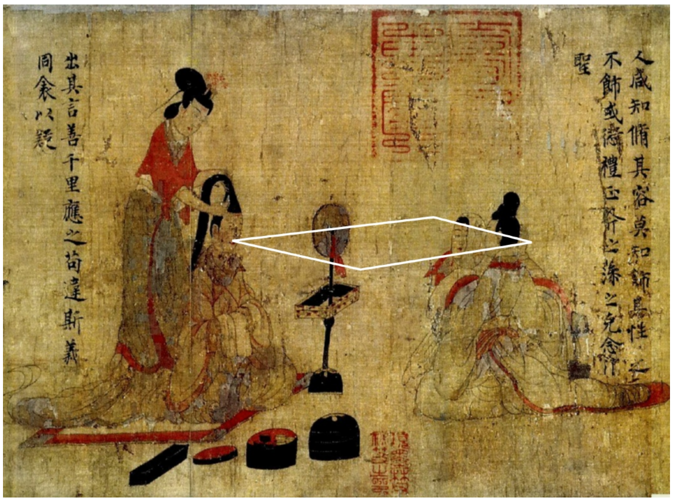

Figure 4.1.4: Parallelogram in the Admonitions of the Court Instructress. British Museum. London

grooming ladies are facing two mirrors.These mirrors imply the existence of two reflections (one has been already drawn, the other is implied to exist). Each lady and her reflection constitute a set of the figure-reflection. Meanwhile, these two sets form another pair of figure-reflection that are reversed to each other. Then, these four images thus form a two-leveled figure-reflection set. (Wu, 1996) The structural relationship of this figure-reflection set is a parallelogram.

In the existing CHSPs, parallelograms are used as the prototypical visual structure to suggest depth. To represent different targets, ancient Chinese painters transformed and combined them, thus creating different visual structural layers of the CHSP.

4.2 Cell-scene image

According to the division of the CHSP, the cell-scene image is the spatial unit. It refers to the projection of a single scene, which consists of one or several object images and a background image.

4.2.1 Visual structure of cell-scene image

The visual structure of the cell-scene image is composed of the visual structure of the object image and the visual structure of the background image.

Object images in the CHSPs have two common visual structures. One is the planar form. Such as the mountain images in the *Recumbent Journey in Xiaoxiang* [Figure 4.2.1-left], whose visual structures are planar triangles. Since a single planar form cannot generate depth, the painter usually staggers and overlaps multiple planar forms to form an oblique image sequence. The diagonal lines connecting the image sequence, like the connection of the mountain triangles, will be perceived as two diagonals of a parallelogram. The other is the composite form consisting of a parallelogram. For example, the building image in the painting of *Spring Morning in the Han Palace*, [Figure 4.2.1-middle]. Its visual structure presents as the projection of a hexahedron that is composed of parallelogram top, bottom, and lateral faces. The visual structure of composite form also appears in the landscape CHSP after the 11th century. In the painting of *Autumn Colors on the Que and Hua Mountains*, [Figure 4.2.1-right] the visual structure of the mountain

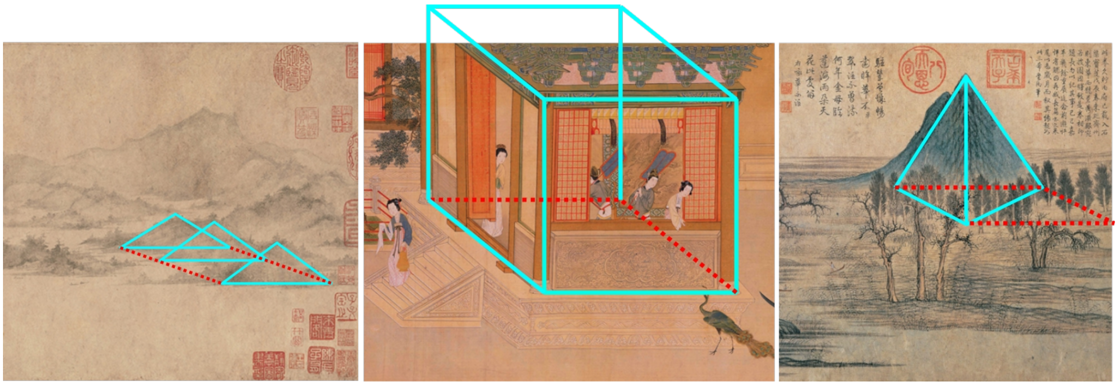

Figure 4.2.1: Visual structure of cell-scene image

Figure 4.2.2: Visual structure of the background image

image shows as a projection of a tetrahedron composed of a triangular base and two triangular laterals. The triangular base is seen as a half parallelogram sharing a common edge.

The background image of the cell-scene image is the remaining image after removing all object images. There are also two types of background images in the common CHSPs. One is the single ground image. It appears mostly in the cell-scene image representing shallow depth. Like the background image in the painting of *Spring Morning in the Han Palace*. [Figure 4.2.2-left] It expands to the top and bottom edges of the cell-scene image and represents a yard's depth. The other is the combination of ground image and sky image. This type of background image is often found in the cell-scene image of a landscape that represents great depth. Such as the background image in the painting of *Dwelling in the Fuchun Mountains* 富春山居图. [Figure 4.2.2-right] It is vertical stitching of the ground image and the sky image. The ground image does not have a noticeable visual structure. It will present to have a visual structure that has the same properties as the visual structure of the object image, only when it is perceived together with the object image. For instance, the ground image in the painting of *Spring Morning in the Han Palace* does not have a noticeable diagonal. However, since it contacts the base of the object image and the visual structure of this base is a parallelogram, this ground image will be perceived as an extension of the parallelogram base which has a structural grid that has the same inclination as the object image's base. Likewise, the sky image is a vertical flat plate without any deformation or distortion。It has no obvious visual structure of its own. But influenced by the front or back of the object image, it will be seen to have a visual structure of an orthogonal grid with no inclination or distortion.

The combination of the visual structures of object images and the background image constitutes the overall visual structure of the cell-scene image. From the whole of the cell-scene image, its visual structure has the following two characteristics. First, the diagonals of all parallelograms

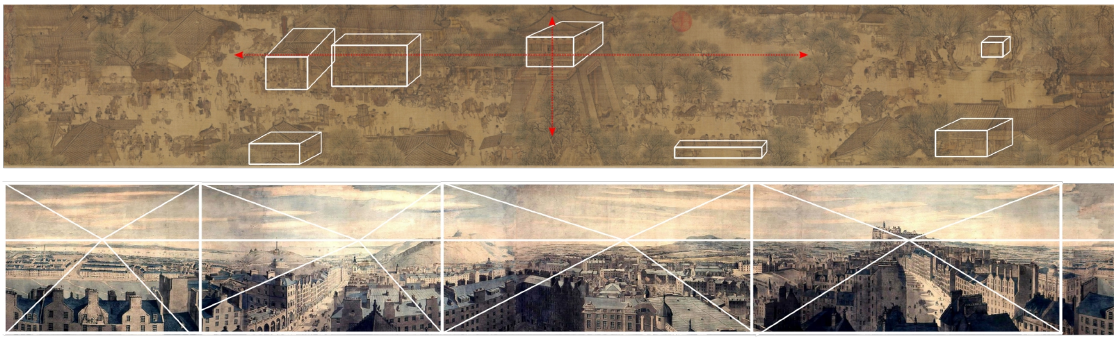

Figure 4.2.3: Compositions of images in the European panorama and CHSP

in the object and ground images have the approximate same inclination. This means that there is no size diminution along the depth direction. Second, there is always a face, such as the front of a building image or the back of a mountain figure, whose visual structure remains parallel to the canvas's surface. These two features allow cell-scene images to be freely combined without structural conflicts. This can be explained by comparing Zhang Zeduan's 张择端CHSP of *Along the River During the Qingming Festival* [Figure 4.2.3-top] to Robert Barker's panoramic painting of "The Panorama" [Figure 4.2.3-bottom]. Both Zhang and Barker use a horizontal combination of scene images to compose their panoramas, but they do produce very different effects. For example, the combined scene images can be easily distinguished from Barker's panorama, but not from Zhang's CHSP. The reason for this difference lies in the different visual structures of the scene images being combined. In Barker's panorama, the visual structure of combined scene images is a one-point perspective, which is characterized by convergence. The depth structure lines of all graphs in such scene image point to a unique center. When multiple scene images with the such visual structure are placed in the same frame, the structural lines of two adjacent scene images will conflict at the junction because of their different convergence position. Therefore, even if they are placed on the same canvas, they are still independent of each other.[1] The front faces of the scene images in Zhang's paintings are also parallel to each other, as are the diagonals in the lateral faces of object images. When they are grouped together, their visual structure will only produce parallel or nested correlation without structural conflicts. From this point of view, this visual structure has the advantage of expanding infinitely in all directions compared to the one-point perspective.

4.2.2 Formation of cell-scene image

In the study of Chinese painting, there are various explanations for the formation of a single scene image. Bachhofer (1931) defines it as a "parallel perspective", as the parallel lines in most Buddhist murals converge toward the center of the image.[2](Tseng, 2006) Fong (1969) also describes it as a "parallel perspective", but he emphasizes that this visual structure is the result of planar operations rather than a perspective based on optical projection in the usual sense. Wu (2004) defines it as "frontal-oblique projection" to describe the coexistence of a parallel front face and oblique lateral faces. Although these definitions have implicitly described the formation of a single scene image, such as perspective, projection, etc., they are conceptual descriptions

[1] This panorama has to be placed on a cylinder wall.

[2] Bachhofer summarizes the characteristics of ancient Chinese art as a trinity of a downcast view, a parallel perspective, and a continuous narrative.

Figure 4.2.4: Woodcut painting of Albrecht Dürer

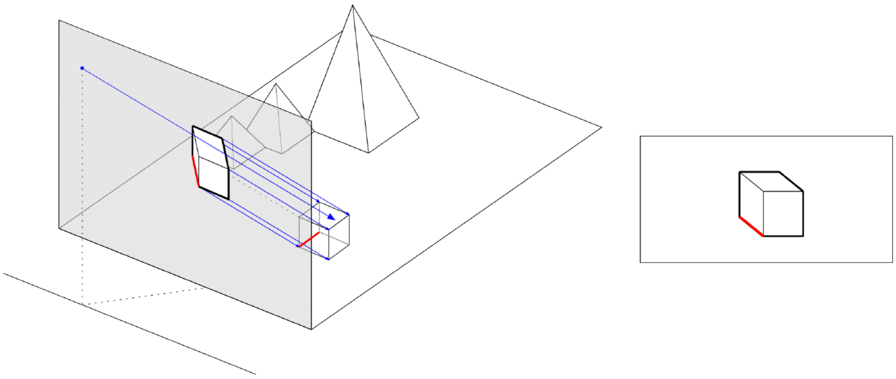

Figure 4.2.5: Composition of cell-scene image

in a specific context and do not address the specific operational process of formation. For this reason, it is still necessary to conduct further research based on the previous.

There are three principles in the ancient theory of painting and architectural representation that are directly related to the formation of a single scene image. The first principle is "stretch the silk to projection the far scenery." [3] which proposed by Zong Bing 宗炳in his article of *Introduction to Painting Landscape* 画山水序. This principle implies the formation of the cell-scene image has the characteristics of optical projection. Combining the fact that there is always an undistorted front in the cell-scene image, it can be inferred that the projection plane (the silk in the painter's hand) must be perpendicular to the ground. Therefore, the optical projection installation that Zong described is similar to that of perspective projection illustrated by Albrecht Dürer in his woodblock prints. [Figure 4.2.4]However, judging from their visual structures, they follow different rules. As the diagonals in the cell-scene image are parallel to other, Zong's projection rule is similar to the rule of oblique projection. [Figure 4.2.5]

The second principle is "Although one point and one stroke must follow the ink rope and the ruler."[4], which was recorded in the *Palace Catalog of Xuanhe Huapu* 宣和画谱of Song Dynasty. This principle means that the size of the object image has to be scaled down proportionally. This principle comes to a severe level in the *Jie Hua* 界画（ruler painting)[5]. For example, the Northern Song scholar Li Zhi 李廌 described Guo Zhongshu's 郭忠恕(c.929 – 977) strict

[3]Original text: 张绢素以远。

[4]Original text: 虽一点一笔，必求诸绳矩。

[5]The *Jie Hua* 界画is a type of painting popular in the Song Dynasty. This painting is drawn by *Jie Chi* 界尺, which is a drawing rule similar to the parallel ruler used by modern architects. It helps the painter draw parallel lines accurately and calculate oblique lines' lengths according to their inclination and actual length.

Figure 4.2.6: Summer Palace of Emperor Ming Huang. Guo Zhongshu. Osaka City Museum of Fine Arts, Osaka

adherence to scale and architectural modular, in his book of *Deyuzhai Huapin* 德隅斋画品. [Figure 4.2.6]

> "Roof beams, girders, pillars, and rafters are shown with open spaces between, through which one might move. Railings, lintels, windows, and doorways look like they could be passed through or opened and shut. He has used a "Hao" 毫to mark off a "Cun" 寸, and use a "Cun" to mark off ten "Chi" 尺; increasing thus with every multiple, so that when he does a large building, everything is to scale and there are no small discrepancies."[6] (Maeda, 1975)

Here "Hao", "Chi" and "Chi" are units of length in the ancient Chinese architectural modular system of *Cai Fen* 材分[7]. Therefore, Guo Zhongzhu's paintings can even be considered as construction drawings.

The third principle is "In painting constructions, calculations should be faultless, and brushstrokes of even strength should deeply penetrate space, receding in a hundred diagonal lines."[8] (Maeda, 1975), which was summed up by Guo Ruoxu 郭若虚of the Song Dynasty in his book of *Shuhua Jianwen Zhi* 书画见闻志. This principle suggests that the recession of space is represented by the diagonal and that the length of the diagonal needs to be strictly proportional. "Receding in a hundred diagonal lines" 一去百斜means that the length of the oblique line needs to be proportional to the depth of the intended expression. "Calculations should be faultless" 折算无亏means that the length of the lateral edge should be reduced with the length and width of the building. (Fu, 2001)

These three principles are the three necessary conditions for the formation of the cell-scene image. Here, I made an assumption that ancient painters strictly adhered to these three principles to paint. Then, when he projects by the first principle, the size of the object he sees from the projection plane is the size of the object that has been reduced by perspective. And because of the concept of representation in painting[9], then it can be inferred that the proportions to be strictly followed in the painter's second principle are the proportions in which the actual object

[6]Original text: 以毫计寸，以分计尺，以尺计丈，增而倍之，以作大宇，皆中规度，曾无小差。

[7]*Cai* is the standard for the length of architectural components in the Song dynasty's architectural manual of *Yi Zao Fa Shi* 营造法式(Liang, 1983)

[8]Original text: 画屋木者，折算无亏，笔画匀壮，深远透空，一去百斜。

[9]See this concept in 2.3

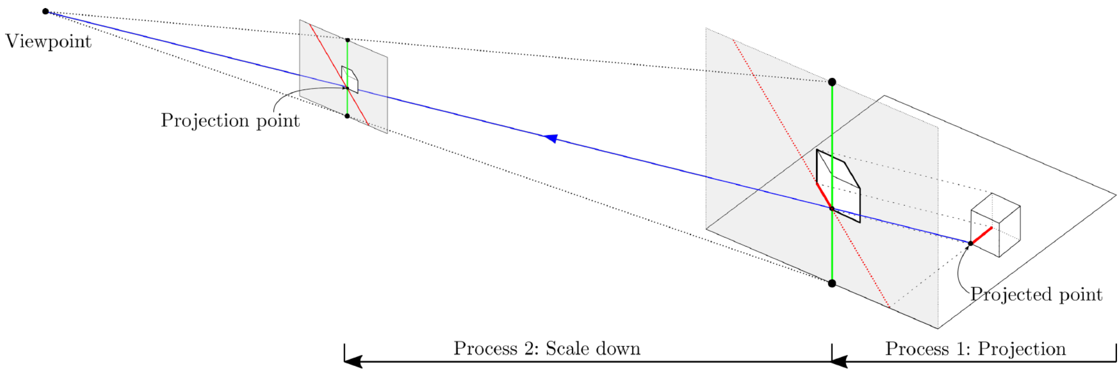

Figure 4.2.7: Drawing method of classical Chinese painting

appears in his view. Meanwhile, since the front face of the object image is not distorted, then all dimensions of this front face are in full compliance with this scale. Also because of the concept of representation in painting, the inclination of the diagonal in the object image must correspond to the direction of his sightline, so as to ensure that the length of the diagonal on the side of the object image is consistent with the length of oblique edge shows of the object. According to the third principle, the diagonals of object images have to be drawn proportionally to this frontal.

Taking all these inferences together, the formation of the cell-scene image is likely to involve a two-step composite process. First, the painter adopts Zong Bing's method to observe the scene and determines the size of the front faces of the projection based on the object's actual size in the field of view. Second, he projects the frontal faces of the objects and then draws the other faces by following the oblique projection rule. In this drawing process, the direction of the painter's sightline is used as the unified direction of the oblique projection. This formation can be seen as a simpler two-step process, as the diagram illustrated in [Figure 4.2.7].The first step is to conduct a regular oblique projection. The second step is to scale down this projection according to the size of the object in the view. The small projection plane in the illustration is the real projection plane in the artist's hand. The large projection plane is the imaginary projection plane that has not been reduced. Both two projection planes are perpendicular to the ground. Since the oblique projection is scaled according to the object's size in the view, then the relationship between the small and large projection planes follows the diminishing rule in perspective projection.

4.3 Column-scene image

The column-scene image is a temporal unit of the CHSP that the viewer sees at a time. It refers to the representation of a scene with great depth. However, the column-scene image is not a complete projection of this scene, but vertical stitching of multiple cell-scene images. In landscape painting, ancient Chinese painters had additional rules for the cell-scene images being combined. Guo Xi's 郭熙indicates in his book of *Lin Quan Gao Zhi* 林泉高致that,

> "The mountain has three distances, look up from bottom to the top, called High Distance 高远; look through from the front to the back, called Deep Distance 深远; overlook from near to the distant, called Level Distance 平远."[10] (Wang, 1981) [Figure 4.3.1]

[10]Original text: 山有三远，自山下而仰山巅，谓之高远；自山前而窥山后，谓之深远；自近山而望远山，谓之平远。

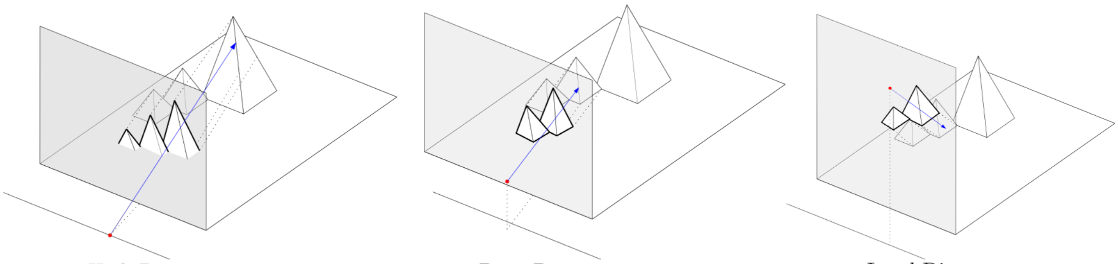

Figure 4.3.1: Three Distances principle of Guo Xi

A complete representation of a mountain should include visions of three distances. [Figure 4.3.2]This means that the artist needs to acquire images of the mountain from at least three angles in three positions and combine them in one frame. In this sense, the purpose of the vertical composition of cell-scene images is to represent the complete space in a limited vertical frame. Li Cheng defines this vertical composition of cell-scene images as "adjusting the height 摆布高低in his monograph of *The Secret of Painting Landscape* 画山水诀.

> "When painting, the painter must first decide the position of the primary and secondary images in the painting; then decide the shape of the objects near and far, and finally, insert and adjust the height of object images."[11]

As can be seen, the formation of a column-scene image is a pure pictorial operation.

Combining the above two arguments, we can infer the formation of the column-scene image. First, the painter needs to obtain cell-scene images from different angles at different positions. Second, he needs to arrange them from top to bottom, according to the depth represented by these cell-scene images. [Figure 4.3.3]

The above inferred is the general formation process of the column-scene image. Along with the development of the CHSP[12], the composition of the column-scene has evolved three times and this resulted in three different ways of compositions: compartmental juxtaposition, partial overlap, and continuous fusion. These compositions make the column-scene image present three different visual structures.

4.3.1 Visual structure of column-scene image in the first stage

The composition of the column-scene image at this stage presents as the vertical juxtaposition of three compartmental cell-scene images. The painting of *Goddess Luo Rhapsody* is one of the typical CHSPs has such composition[13].

In the selected column-scene image of this CHSP [Figure 4.3.4], the object images are vertically separated into three groups by blank spaces: mountains in the upper segment, men in the middle segment, and waterfront in the lower segment.The visual structures of the figure and tree images are flat forms, while the other object images have the visual structures of composite forms. In these object images, the base of the mountain, the side of the drum, and the texture of the river image all imply diagonal lines, as indicated by the bolded diagonals in [Figure 4.3.4].They

[11]Original text: "凡画山水，先立宾主之位，决定远近之形，然后穿凿景物，摆布高低。"

[12]See the division of three-staged development of the CHSP in 1.3.2

[13]The edition of *Goddess Luo Rhapsody* in Beijing Palace Museum is supposed to be a copy made in the late Northern Song Dynasty (c. early 12th century). This version is thought to preserve the original composition of the Six Dynasties period (Chen, 2011).

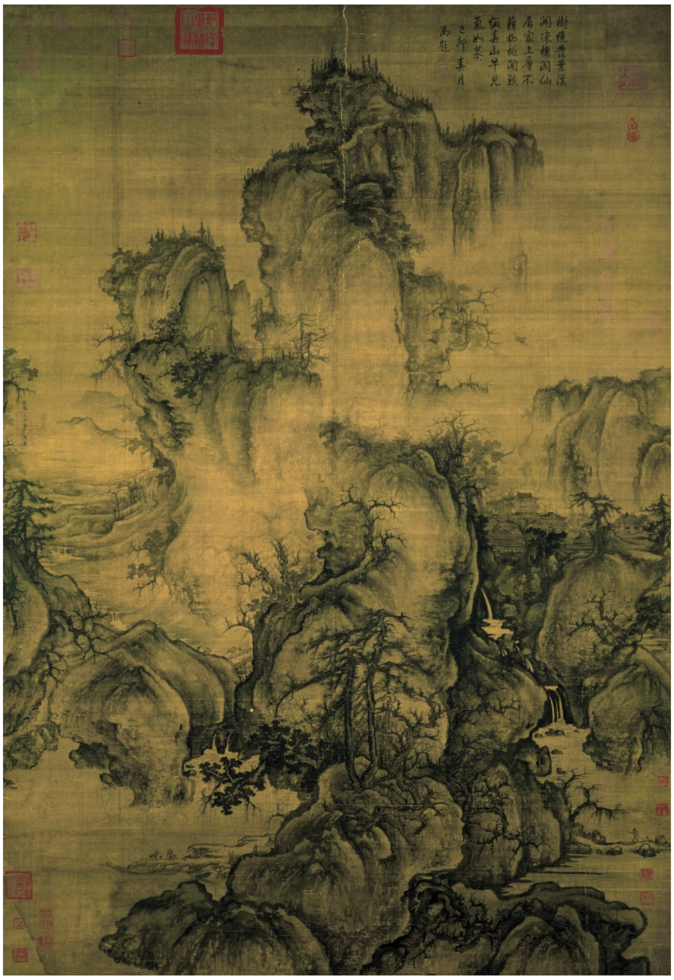

Figure 4.3.2: Early Spring. Guo Xi. Taipei Palace Museum, Taipei

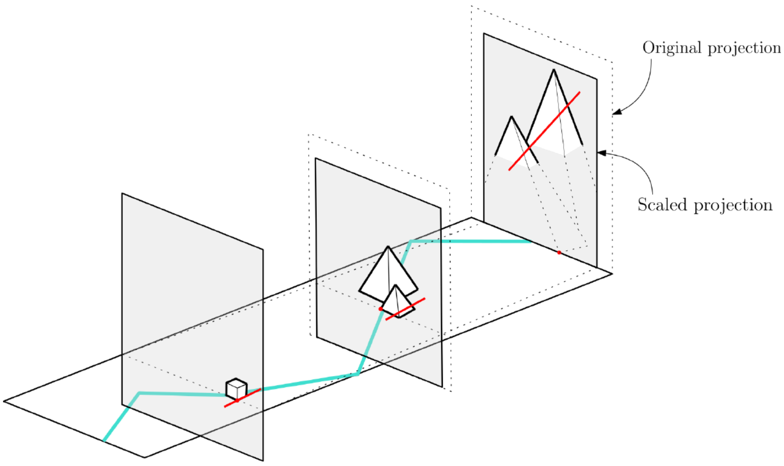

Figure 4.3.3: Composition of the column-scene image

Figure 4.3.4: Composition of the column-scene image in *Goddess Luo Rhapsody*

imply the depths in their respective segments, and they have different inclinations. At the same time, all three groups of object images are reduced in size, but to different degrees. This suggests that the artist chose different size reduction ratios when depicting them, depending on the size of their appearance in the view.

This ground image itself has no noticeable visual structure. According to the analysis of the visual structure of the background image in 4.2.1, the ground image can be considered as a larger parallelogram that has the same properties (inclination and size reduction ratio) as the parallelogram in the base of the object image it bears. Therefore, the visual structures of the ground image in the three segments are three grid structures with three inclinations and three different densities. The visual structure of the whole ground image of the column-scene image shows a vertical collocation of three grid structures. [Figure 4.3.4]

4.3.2 Visual structure of column-scene image in the second stage

The composition of the column-scene images at this stage is still the vertical composition of multiple cell-scene images. But different from that of the CHSP in the previous stage, more cell-scene images are composed and partial overlaps appear between some neighboring cell-scene images. These features are most evident in the painting of *Recumbent Journey in Xiaoxiang.* [Figure 4.3.5]

The selected column-scene images show that the mountain images appear in groups and are very closely combined. In each group, the visual structure of a single mountain image is presented in a planar form. And they are staggered to form a tilted form sequence. The lines connecting the bottom edges of the mountains in each sequence form a parallelogram, as indicated by the bolded diagonals in [Figure 4.3.5] It suggests the depth of the area in which the sequence is located.

In different groups of mountain images, the diagonals in the form sequence have different slopes and also show different degrees of size reductions. Same with other cell-scene images, these mountain images' visual structures determine the visual structures of their respective ground images, making them parallelogram grids with the same inclinations and size reduction ratios as the bases of the objects they bear.

These features are not fundamentally different from the visual structure of the column-scene image in the first stage. However, partial overlap occurs between some adjacent cell-scene images, which causes a qualitative change in the visual structure of the ground image in the overlapping

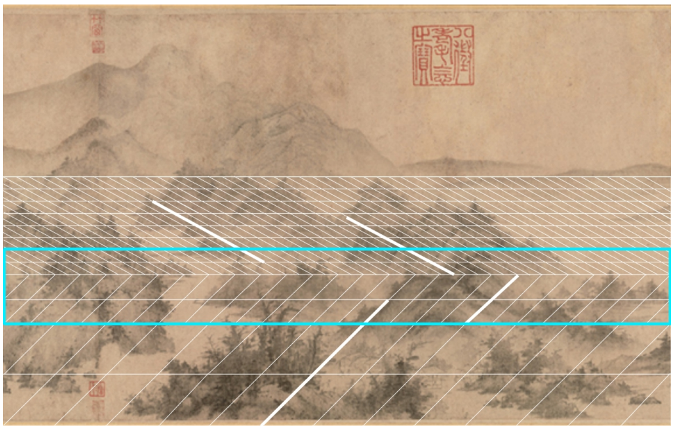

Figure 4.3.5: Composition of the column-scene image in *Recumbent Journey in Xiaoxiang*

region. For example in [Figure 4.3.5], most of the mountain image groups are isolated from each other. But in the boxed section, the lower group of mountain images obscures the upper one. Since the visual structure of the ground image is determined by the visual structure of the object image it is carrying, then the partial overlapping of the ground image means that two kinds of visual structures appear in this overlapping area. This mixture of visual structures would create a polysemy in the visual sense. One interpretation is that the visual structure of this part of this overlapping region would be perceived as the extension of the visual structure of the upper or lower cell-scene image. The other explanation is that it would be perceived as a transitional visual structure between the upper and the lower cell-scene images. In this case, the inclination of the diagonals and the size reduction ratios of the images in the overlapping area would change gradually between the upper and lower cell-scene images.

In summary, the visual structure of the entire ground image of the column-scene image is generally stitched of multiple parallelogram grids. However, in the same overlapped parts, these grids are changing gradually between the overlapped cell-scene images. [Figure 4.3.5]

4.3.3 Visual structure of column-scene image in the third stage

The most remarkable change in the column-scene image at this stage is that a composite of object images that runs through vertical directions, such as the image of rolling hills or zigzag riverbanks, appears in the composition. This composition makes the column-scene image present as a coherent whole. Huang Gongwang 黄公望(1269–1354) defined this composition as Level Distance 平远in his painting treatise of *Shanshui Jue* 山水诀. The meaning of Level Distance is that (the mountains) continuously connected from near to far is Level Distance.[14] (Wang, 1981) His painting of *Dwelling in the Fuchun Mountains* is an example of such composition. [Figure 4.3.6]

As seen in the selected image, the mountain images in this image no longer have any blank space in between but continuously joint head to tail to form a coherent composite image. This composite image extends uninterruptedly from the bottom up to the top of the column-scene image. In this composite image, the sizes of the combined mountain images are reduced differently. Meanwhile, their visual structures show as composite forms of the tetrahedron whose base

[14]Huang's definitions of Three Distances: (the mountains) constantly joined from below, is called Level Distance. Separated from each other from near to far, is called Wide Distance. A distant view from beyond the mountain is called High Distance. 从下相连不断。谓之平远。从近隔开相对。谓之阔远。从山外远景。谓之高远。

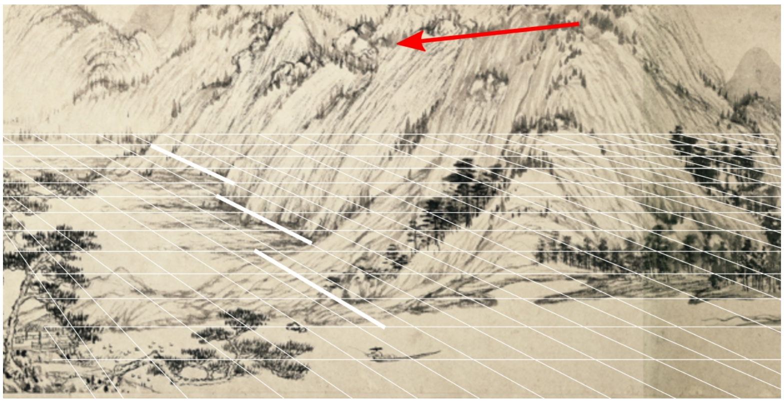

Figure 4.3.6: Composition of the column-scene image in *Dwelling in the Fuchun Mountains*

can be seen as a half parallelogram. And the diagonals in these parallelograms have different inclinations. Based on the analysis of the visual structure of the background image in 4.2.1, the visual structures of the composed mountain images determine the visual structures of the water images on which they are located. These ground images will then be considered as larger parallelogram grids with the same inclinations and size reduction ratios as their respective mountain images.

From the above analysis, it can be seen that the visual structure of the column-scene image in this stage is not fundamentally different from that in the other two stages. However, if we carefully compare the size reduction and diagonal inclination between the combined mountain images, it is easy to see that they both show a gradual change. The change of size reduction ratio is most evident in the area delineated by the red arrow. The ridges in this area even present a nearly smooth diagonal line. This phenomenon would create an effect in the visual sense that the combined single mountain images' sizes are reduced with a very small difference each time. As each size reduction ratio corresponds to a single mountain image, this continuous change means that there are countless hidden mountain images being inserted among the obvious mountain images. With this visual implication, numerous hidden diagonals are then considered to be inserted between any two diagonals with a large difference in inclination, and their inclination varies continuously between the inclinations of these two diagonals.

Based on the above visual senses, the visual structure of the ground image of the entire column-scene image will be perceived as a vertical composition of numerous parallelogram grids with their size reduction ratios and inclination change continuously. When the size difference between two adjacent ground images is tiny, the densities in their grids would not change jump-like but change continuously. In this case, the connection of the grids along the vertical direction will not appear like a stacked arrangement of parallelograms in the [Figure 4.3.5], but like smooth connections of trapezoids [Figure 4.3.6].

Compared with the features of the visual structures of the column-scene images in the three stages, a trend in the development of the column-scene images can be identified. That is, more and more cell-scene images are being combined, the differences between their visual structures are becoming smaller and the transitions are becoming more continuous. This makes the column-scene image they combined increasingly coherent. Thus, it can be concluded that the purpose of the evolution of the column-scene image is to represent a coherent scene with great depth.

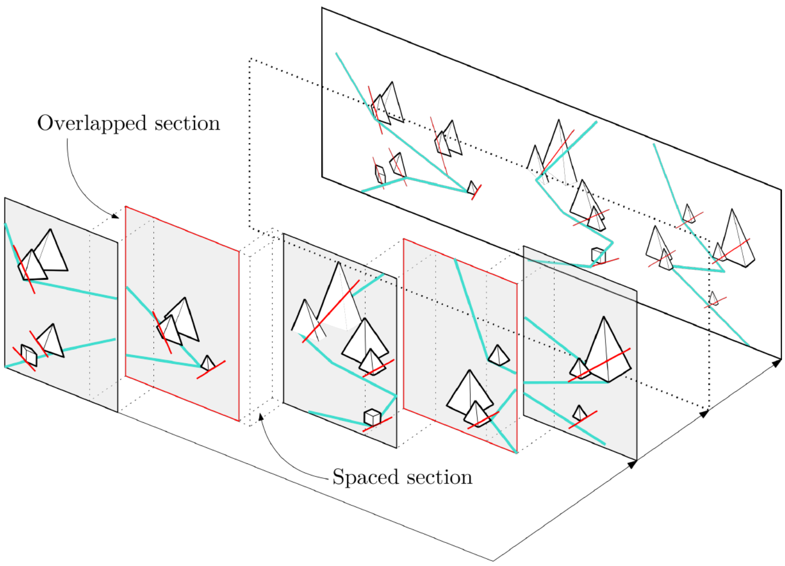

Figure 4.4.1: Composition of the row-scene image

4.4 Row-scene image

The purpose of the row-scene image is to complete the narrative of a series of scene images. Thus, in the formation of the row-scene image, the painter needs to deal with two necessary issues: the arrangement of column-scene images and the transition between column-scene images.

According to the analysis of the scene images in the painting *Autumn Colors on the Que and Hua Mountains* in 1.3.2, the row-scene image is a horizontal combination of column-scene images obtained from several different positions and at different angles. From this, it can be deduced that the composition process of the row-scene image generally consists of steps. First, arrange horizontally the column-scene images in a certain sequence according to the narrative of the story or the path of the journey. Second, process the transition and articulation between the column-scene images. [Figure 4.4.1].

During the development of CHSP, the ancient painters' treatment of the transition between column-scenes went through three-staged changes and resulted in three different ways of compositions: compartmental juxtaposition, partial overlap, and continuous fusion. These compositions make the column-scene image present three different visual structures.

4.4.1 Visual structure of row-scene image in the first stage

The row-scene image at this stage follows strictly the early compositional convention of *left picture and right text* 左图右史[15] which means putting vertical textual annotation to the right side of the image. In this composition, the vertical text isolates two adjacent column-scene images as a boundary, making the row-scene image present as a horizontal juxtaposition of compartmental column-scene images.

This combination method is mainly applied in early CHSPs. Such as the painting of *Admonitions of the Court Instructress*. The image in [Figure 4.4.2] shows a selected fragment of this painting that is composed of two column-scene images. The right one shows a scene in which

[15]Xue believes that format of left picture and right text is the prototype of the CHSP (Xue, 1993).

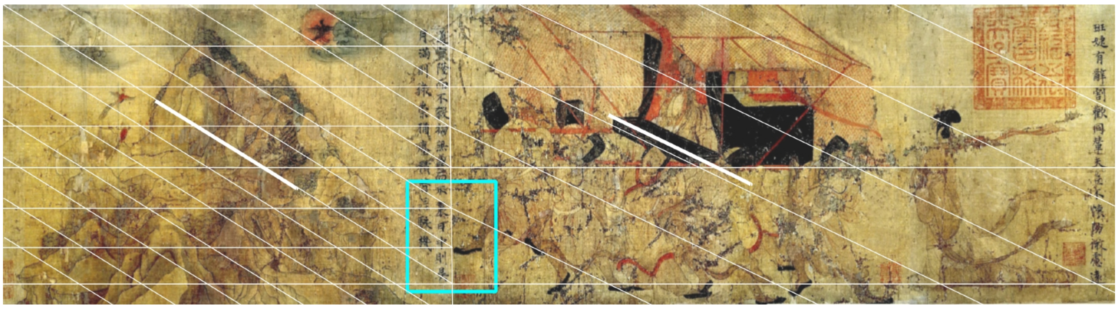

Figure 4.4.2: Composition of the row-scene image in the *Admonitions of the Court Instructress*

palanquin bearers carry the Emperor Chengdi 汉成帝of Han on a trip. The left depicts a landscape where a hunter is hunting. As can be judged from the lateral waistline of the mountain image on the left and the lateral edge of the car on the right, the diagonals implying spatial depth in the visual structures of these object images have different inclinations. Meanwhile, the differences between their sizes indicate that they were scaled down to different degrees.

According to the relationship between the visual structure of the background image and the visual structure of the object image, the visual structure of the ground image would be perceived as a structural grid, which has the same inclination and size reduction ratio as the object image it bears. Therefore, the differences in the inclination and the size reduction of the object images mean that the inclinations and densities of the structural grids of the ground images of the two column-scene images are different. And due to the segmentation of the vertical text, these structural grids are directly stitched together at their junctions. The visual structure of the ground image of the entire row-scene image is the direct stitch of multiple structure grids with different inclinations and densities.

Besides the above direct juxtaposition of the column-scene images, there is a variant of compartmental composition. This variation was applied in Gu Kaizhi's other painting of *Goddess Luo Rhapsody*. As the selected fragment shows in [Figure 4.4.3], the cell-scene images in this composition break the boundaries of the column-scene image and penetrate into the neighboring column-scene image. This penetration implies a spatial fusion between adjacent column scene images. However, the remaining vertical texts between cell-scene images indicate that this fusion is limited at the column-scene image level, the visual structure of these cell-scene images still remains closed and independent of each other. Accordingly, the visual structure of the whole ground images is still a direct stitch of multiple structure grids with different inclinations and densities.

4.4.2 Visual structure of row-scene image in the second stage

The composition of the row-scene image at this stage abandoned the convention of compartmental juxtaposition with vertical texts and started to partially overlap the adjacent column-scene images. This composition then creates a shared region in the overlapping area. This shared area means that the boundaries between these column-scene images are blurred and the internal spaces of these column-scene images are fused.

Gu Hongzhong's 顾闳中painting of *The Night Entertainments of Han Xizai* 韩熙载夜宴图is one of the typical examples have this composition. In the selected image in [Figure 4.4.4], the men image wearing a tall hat are the same person, Han Xizai. As he appears on both sides of the screen simultaneously, it can be inferred that the scenes on both sides of the screen occur at

Figure 4.4.3: Composition of the row-scene image in *Goddess Luo Rhapsody*

different times during the banquet. Then, theoretically, these two scenes should be bounded by the screen[16], isolated and independent of each other. Their depths are implied by the diagonals in the visual structures of the chair image on the left and the visual structure of the screen image on the right. Judging from the inclinations of the diagonals, the visual structures of the chair image and the screen image are different. This difference also further affects the ground images on both sides of the screen, making their visual structure perceived as two structural grids with different inclinations. Therefore, the visual structure of the whole ground in this row-scene image is a stitching of two structural grids with different inclinations.

However, the appearances of the screen and the figures on either side of it present negation of the above conclusion in terms of spatial experience. First, the scenes on both sides of the screen are not bounded by the screen in terms of content. This can be proved by the two people on either side of the screen looking at each other. This phenomenon shows that they are communicating with each other and that they belong to each other's scenes. Therefore, a more reasonable way to divide these two column-scene images is to group these two characters into each other's scenes, as shown in [Figure 4.4.4]. This division means that two adjacent column-scene images are partially overlapped, producing a shared area. In the overlapping area, the visual structure of the ground image will present an uncertain state: it can be an extension of the visual structure of the ground image on the left, or of that on the right, or a structural transition between the two. Second, the two scenes are spatially coherent in pictorial space, although they are temporally independent of each other. This judgment is due to the nature of the screen. Unlike a normal wall, the screen does not touch the roof or even the walls on either side. In other words, the screen is only a partial partition of the space. The spaces on either side of it are a continuum, including the floor, as the blue arrow indicates in [Figure 4.4.4]. The coherence of the ground image means that the visual structure in the shared region would not be perceived as an extension of the left or the right, but more likely to be a continuous transitional structure. That is to say, the inclination of the structural grid of the ground image in this region is more likely to change gradually from that of the left to that of the right, which presents as swaying diagonals in the shared region.

The composition analyzed above is also applied in the CHSP of *Vanquishing Mara* 降魔图in the Dunhuang Caves. The selected image in [Figure 4.4.5] shows centripetal scenes in this scroll painting.In the scene on the left, the men are watching an elephant; in the scene on the right,

[16]Screen 屏风is the special furniture to divide the interior to partition a room in Chinese architecture. It is usually painted with calligraphy or paintings.

Figure 4.4.4: Composition of the row-scene image in *The Night Entertainments of Han Xizai*

Figure 4.4.5: Composition of the row-scene image in the scroll of *Vanquishing Mara*. Bibliothèque nationale de France, Paris

the men are watching a lion fighting a bull. The two scenes are bounded by a tree. However, in both scene images, there appears a person looking outward, such as the woman and the monk in the boxes. These figures seem to be more properly grouped into their adjacent scenes in terms of content. This means that the range of these two adjacent column-scene images is changed and a shared area is created between them that belongs to both scenes at the same time. Also, because the tree between the two scene images is not completely separated from the ground images on both sides, the ground images in the shared region are coherent. Then the visual structure of the ground images in the shared region will be perceived as a structural grid whose slope varies gradually between the slopes of the left and right scene images, as shown in [Figure 4.4.5]

4.4.3 Visual structure of row-scene image in the third stage

At this stage, the composition of the row-scene image no longer has a distinct stitching feature. All the column-scene images are fused together and cannot be easily distinguished. This composition is most common in *Jie Hua*[17]. Such as Qiu Ying's 仇英painting of *Spring Morning in the Han Palace. Qiu Ying. Taipei Palace Museum, Taipei*汉宫春晓图. The image in [Figure 4.4.6] shows a part of the row-scene image of this painting.This fragmented row-scene image has a continuous building façade as its background. A platform image with a parallelogram base and parallel flower bed images are arranged in this image. The diagonals in the visual structure of these images suggest the depth of the entire image. These diagonals appear to be parallel.

[17]Jie Hua or the Ruler Painting is a type of Chinese painting which is drawn with rulers. These paintings are usually themed on a palace or city.

Figure 4.4.6: Composition of the row-scene image in *Spring Morning in the Han Palace*

Figure 4.4.7: Jie Chi used by Zhang Zheduan, Exhibition of River and Wisdom

Therefore, this fragmented row-scene image looks more like a larger cell-scene image, which is gained by the projection method in 4.2.2.

However, if we compare the inclinations of the diagonals on both sides of the platform, we can see that there are some slight differences between them, as shown by the solid black diagonal and the solid white diagonal on the far left in [Figure 4.4 6].In some studies (Osborne, 1970), this phenomenon was treated as a kind of error due to the lack of rigor or scientific method. However, the *Jie Chi* used by ancient painters, which helped them to draw precise parallel lines, proves that the possibility of such a situation is extremely low. [Figure 4.4.7]If we combine the previous analysis of the composition of the row scene image with the analysis of the projection rules in 4.2.2, we can see that the difference is more likely due to the painter changing the projection direction.[18] This assumption can be proved by two groups of staggering figures (boxed) in the left section of this row-scene image: two people lying in bed indoors and two women watching others play outdoors.Judging from the staggered relationship between the front and back figures, the painter projected the two groups of figures at a slight angle off the front, so the back figures are much obscured by the front ones. This projection direction is rough as shown by the yellow arrow. According to the geometric relationship between the projection ray and the diagonal, the projection direction (yellow arrow) corresponding to the white solid line is more off-vertical than the projection direction (red arrow) corresponding to the black solid line. Then, if the painter followed the projection's direction (red arrow) corresponding to the solid black line, the back figure in the frame would be more obscured by the front figure. Therefore, to portray more details of the figure behind, the painter needed to deflect the projection direction to the left when

[18]Tyler and Chen (2011) argues that this deviation in the parallel lines is to correct distortions in visual perception.

acquiring the projection of the scene on the left, which resulted in a deflection of the inclination of the diagonal in the visual structure of the left column-scene image.

However, what can be seen is that this difference is not particularly significant. This reason is that the painter gradually deflected the diagonals between two diagonals with obvious differences in inclination. This can be proved by stroking out all the diagonals (dashed lines) on the platform, such as the edges of the column bases and tables. It can be seen that these diagonals show a swaying effect, with a slight difference between the two adjacent ones. Relating to the "boundary ruler" used by the painter, he was able to draw absolutely precise parallel lines. Then this phenomenon of gradual deflection was most likely his intentional creation. This operation is a bit like the "painterly" technique used by Baroque painters to blur the border between two scenes so as to create a transition between them. (Wolfflin, 1929) In this way he wanted to create a coherent structural transition between two diagonals with significantly different inclinations, thus weakening the differences between them.

The painter's adjustment of the diagonals has a decisive influence on the visual structure of the row-scene image. As explained by the reification principle of Gestalt, while watching the swaying diagonals, the viewer will fill in more gradually rotated diagonals in empty areas, such as the empty space between two columns bases, to create a coherent visual effect.[19](blue short diagonals) [Figure 4.4.6]This will create an illusion that there could be numerous diagonals with gradually changing inclinations between any two adjacent obvious diagonals.[20] According to the analysis in 4.2.1, an inclination corresponds to a cell-scene image, the diagonals with numerous inclinations mean that numerous cell-scene images are being juxtaposed between the two column-scene images. Since there is no combination of cell-scene images in the vertical direction, the row-scene image is essentially a continuous composition of numerous cell-scene images. Accordingly, the visual structure of the entire ground image is a continuous composition of structural grids with continuous inclinations.

This composition is also applied in the CHSP of landscape, like the painting of *Shanglin* 上林图of the Ming Dynasty. [Figure 4.4.8] In the selected row-scene image, the inclination of the diagonal in the left building image and the inclination of the diagonal in the right building image appear to be significantly different. The painter also set the diagonals in the visual structures of the object images located between the two column-scenes to gradually "sway" between the diagonals in the left and right column-scene images as a way to create a transition between them. Meanwhile, he set blank spaces, such as water to separate them. This space allows the viewer to conceive more transitional diagonals with gradual changes between any two sets of diagonals. The row-scene image is then essentially a continuous composition of the visual structures of numerous cell-scene images. Accordingly, the visual structure of the entire ground image is a continuous composition of structural grids with continuous inclinations. However, different from the painting of *Spring Morning in the Han Palace*, the sizes of the object images in the painting of *Shanglin* are reduced along the vertical direction. Moreover, the size reduction ratios of object images of the same height are different in different column-scene images. For example, the sizes of the building images on the blue dashed horizontal line in [Figure 4.4.8]. According to the relationship between the visual structure of the object image and the ground image, this size difference would also determine the density of the structural grid of the ground image. Thus, the difference in the sizes of the object images would cause the densities of the structural grids of the ground images of the row-scene image to be different at the same height.

By combining the visual structure characteristics of the row-scene images in the above three

[19]Reification is the constructive or generative aspect of perception, by which the experienced percept contains more explicit spatial information than the sensory stimulus on which it is based.

[20]The number of diagonals depends on how many divisions the viewer is willing to make of this blank area. In practice, the most extreme case is that each pixel point in the horizontal direction corresponds to a diagonal.

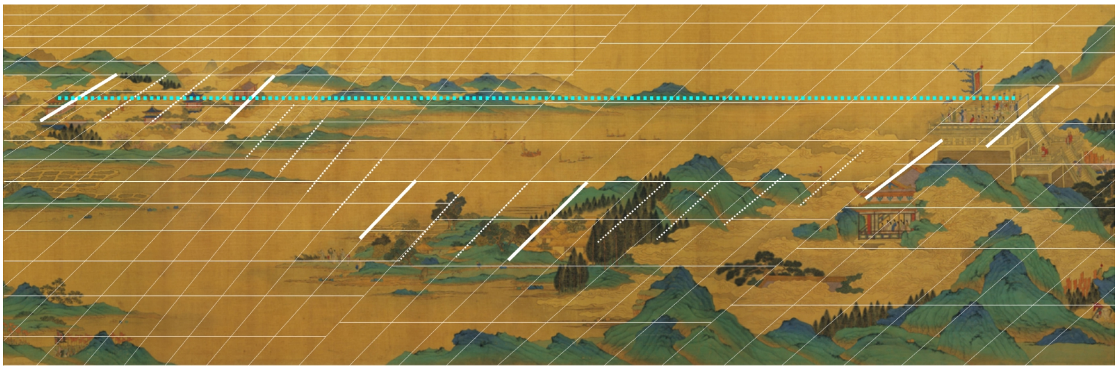

Figure 4.4.8: Composition of the row-scene image in *Shanglin*. Qiu Ying. Taipei Palace Museum, Taipei

stages, it can be found that there was an evolution in the development of the row-scene images. That is, more and more column-scene images are being combined, the differences between their visual structures are getting smaller and the transitions are getting more and more continuous. Therefore, it can be concluded that the purpose of the evolution of the column-scene image is to coherently narrate numerous column-scene images.

Summary

In this chapter, we selected representative CHSPs from different historical stages to analyze the visual structure of the CHSP. Ancient Chinese painters used parallelograms to imply depth recession and compose various scene images: the cell-scene image, the column-scene image, and the row-scene images.

Visual structure of cell-scene image

The cell-scene image is a combination of the object image and the background image. The visual structures of the object image have two types. One is flat form. The other is a composite form with a parallelogram (or half parallelogram) base. The ground image is usually perceived to have the visual structure of a parallelogram grid. The visual structure of a sky image is perceived as a rectangle grid without tilt or distortion. The visual structure of the cell-scene image is the combination of the visual structures of the object image and the background image. It has two characteristics. First, the diagonals in the visual structures of all the object images are parallel. Second, the object image always has a face, or a horizontal structural skeleton remains parallel to the canvas. The formation of object images is inferred as a two-step process. First, the painter adopts Zong Bing's method to observe the scene and determines the size of the front faces of the projection based on the object's actual size in the field of view. Second, he projects the frontal faces of the objects and then draws the other faces by following the oblique projection rule. In this drawing process, the direction of the painter's sightline is used as the unified direction of the oblique projection.

Visual structure of column-scene image

A column-scene image is a vertical composition of cell-scene images. The development of column-scene images aims to represent the scene with large depth coherently.

The composition of the column-scene images in the first stage is presented as a vertical juxtaposition of three cell-scene images. In each cell-scene image, the visual structure of the object image shows as a planar form. The connections of their bottom edges form a parallelogram base. The ground image is perceived to have a visual structure of parallelogram grids that have the same inclinations as the parallelogram base of the object images placed on them. The visual structure of the whole ground image of the column-scene is a vertical collocation of grid structures with three different inclinations and densities.

The composition of the column-scene images in the second stage is still a vertical composition of cell-scene images, but some adjacent cell-scene images are partially overlapped. In different cell-scene images, the parallelogram bases of object images have different inclinations. The sizes of object images in the combined cell-scene images were reduced differently. These inclinations and size reduction ratios determine the inclinations and densities of the structural grids of the ground images on which these object images are placed. In the overlapping area of two adjacent cell-scene images, the visual structure of the ground image is perceived to transit from one visual structure to another, which is reflected in the continuous variation in the inclination and density of the grid structure.

In the column-scene image of the third stage, a composite object image appeared, consisting of multiple object images connecting from head to tail. In this composite object image, the sizes of the combined object images are continuously reduced. As each size reduction ration corresponds to a single object image, the result of this continuously diminishing creates a perception that there compose countless single object images in the composite image with their size reduction rations changing a tiny difference a time. The column-scene image at this stage is then a continuous fusion of numerous cell-scene images. The visual structure of the ground image is determined by the visual structure of the composite object image, which is perceived as a structural grid with continuously varying inclination and density. Since the size difference between two adjacent composed object images is tiny, the change in density of the structural grid is not jumpy but continuous and evenly.

Visual structure of row-scene image

A row-scene image is a horizontal composition of column-scene images. The development of row-scene images aims to narrate the scenes continuously.

In the first stage, the composition of row-scene images is compartmental juxtaposition horizontally, which demonstrates in two ways. One is the juxtaposition with aligned boundaries of vertical texts, in which each of the juxtaposed column-scene images has an isolated visual structure. The other is the juxtaposition with staggered boundaries of vertical texts. In this combination, cell-scene images at different heights start to cross the borders of column-scene images and insert into the adjacent column-scene images. Since the visual structures of these cell-scene images are still isolated, the composition of the row-scene image is still compartmental juxtaposition. The object images in the composed column-scene images have different inclinations and size reduction ratios. These visual structures determine the inclinations and densities of the structural grids of the ground images on which they are located. Then, the visual structure of the entire ground image presents a compartmental juxtaposition of multiple structural grids with different inclinations and densities.

In the row-scene images of the second stage, a shared region belonging to both two adjacent column-scene images is created. These phenomena imply that the two adjacent column-scene images are partly overlapped. There are two representative ways to create a such shared region. One is to use a specific object image motif that can suggest the overlap of adjacent scenes, like the screen. The other is to rotate a figure to make it belong to one column-scene image in semantical but belong to the other in formal. On either side of the overlapping area, the visual structures of the object images have different inclinations and size reduction ratios. These structural features influence the visual structure of their ground images. The structural grids of ground images then to be perceived to have the same inclinations and densities as these object images. In the overlapping area, the ground images of the two column-scene images are fused. The visual structure of the ground images is perceived as a transitional structure between the visual structures of the adjacent ground images, which change continuously in inclination and density.

In row-scene images of the third stage, the diagonals in the visual structures of the object images show a continuous swaying effect in the horizontal direction. This effect creates the illusion that there are countless diagonals with gradually changing inclinations between any two prominent diagonals. According to the fact that each inclination corresponds to a cell-scene image, these countless diagonals mean that countless cell-scene images are composed between any two column-scene images with their inclinations and size reduction ratios varying continuously. Accordingly, the visual structure of the entire ground image is a continuous composition of structural grids with continuous inclinations.

Chapter 5

Visual structure and diegetic movement

This chapter will discuss the geometric correlation between the visual structure of the CHSP and diegetic movement. First, the driving effect of the primordial form in the visual structure of the CHSP in diegetic movement will be explained from the perspective of visual perception. Second, this chapter will elaborate on the inference of the geometric correlation between the visual structure of the cell-scene image and the orientation and position of the diegetic viewpoint. Last, this chapter discusses the effect of the visual structures of the CHSP's column-scene image and row-scene image on the movement of the viewer's diegetic viewpoint based on examples from three development stages.

5.1 Parallelogram and diegetic movement

When we look at an image, the visual system would react according to the relationship between the visual structure of the figure being focused on and its background framework. According to Arnheim (1954), this background framework comes from three sources. The first is the retina, which provides a 2D orthogonal structure. The second is the kinesthetic system that humans inherently have, such as the perception of the direction of gravity. It provides a 3D frame system of "front-back, up-down, left-right". The third one is the visual structure of the background image of the figure.

Visual psychology research (Arnheim, 1954) suggests that the processing of figures is a field process. In this process, the visual system is always in pursuit of the simplest visual structure of a figure or the structural relationship between it and the background frame, due to the law of simplicity. That is to say if Figure A in an image can be seen as another Figure B in Situation B, and either the visual structure of Figure B or the structural relationship between Figure B and the background framework is more concise than Figure A or the structural relationship between Figure A and the background framework, then the visual system will try to "convince" itself to believe in the Situation B. The result of "convincing" is to generate a set of perceptual forces in the perceptual field. This force acts on the figure, trying to transform it to achieve the simplest visual structure or the simplest relationship between it and the background framework. In turn, the counterforce of this force would be generated, which acts on the perceptual system, driving the viewer to move toward a position or shift to a posture, so that the simplest situation can be realized.

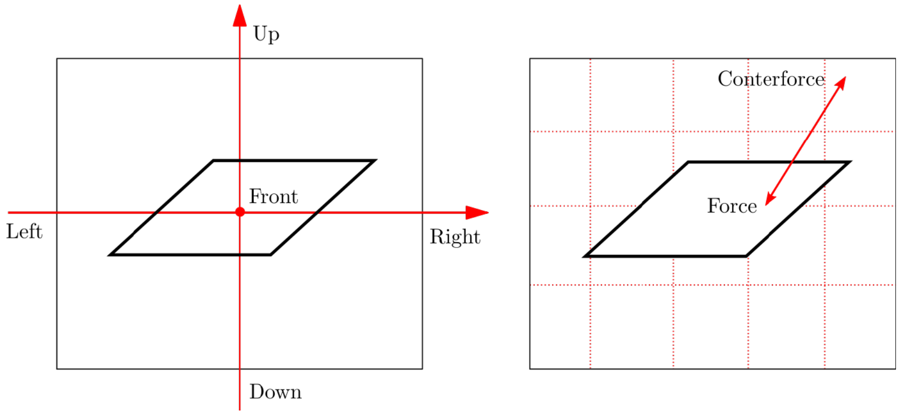

Figure 5.1.1: Parallelogram and perceptual force

When watching a CHSP, this pair of perceptual forces would arise as well because of the conflict between the parallelogram and its background framework. The image of [Figure 5.1.1-left] illustrates the prototype of the structural relation between the parallelogram and its background framework in the CHSP. In this prototype, the background frame is provided by three structures: a 2D visual structure of “up-down, left-right” of the background image, a 2D orthogonal structure of the retina, and a 3D frame system of “front-back, up-down, left-right” of human kinetic system. The three structures overlap, forming a 3D orthogonal background frame whose front-to-back direction is perpendicular to the canvas. Then, the prototype can be interpreted in two ways. One is to see it as a parallelogram in a 2D orthogonal background framework. In this case, the viewer’s sightline is perpendicular to this parallelogram. The viewpoint is located at the front of this image. The other is to see it as a rectangle placed on the ground in a 3D orthogonal background framework. In this case, the viewer’s sightline is tilted. He looks at this rectangle from a right back position in space towards the lower left. Obviously, both the visual structure of the rectangle itself and its structural relationship with the 3D orthogonal background framework are clearly simpler than the parallelogram and its structural relationship with the 2D orthogonal background framework. Therefore, this parallelogram is most likely to be seen as the second case because of the law of simplicity. Once the visual system accepts this setting, this parallelogram would conflict with the 3D orthogonal background framework. This results in a set of perceptual forces in the perceptual field to try to achieve this simplest configuration. The force attempts to push the parallelogram onto the ground. The counterforce subjects the viewer’s visual system and drive him to move towards the right-rear and look down to the left, so that the rectangle on the ground can be seen as this parallelogram. [Figure 5.1.1-right]It can be seen that the direction of this movement is related to the inclination of the parallelogram. In this sense, the parallelogram provides the dynamics and the direction of the diegetic movement.

5.2 Visual structure of cell-scene image and direction and position of diegetic viewpoint

5.2.1 Perceptual force and visual structure of cell-scene image

According to the above analysis, the background framework in the visual field will be determined by three structures when the viewer is watching CHSP. The first is the rectangular outer frame

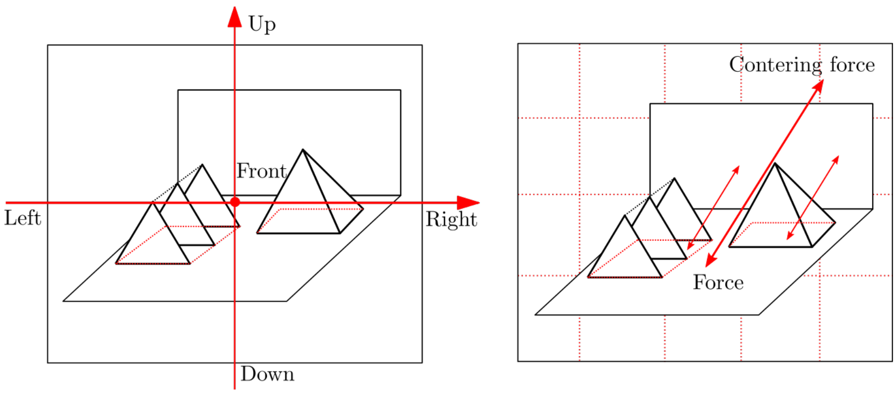

Figure 5.2.1: Visual structure of the object image and perceptual force

of the unrolled frame which has a 2D orthogonal frame system. The second is the 2D orthogonal structure of the retina. The third is the human inherent 3D frame system of "front-back, up-down, left-right". These three frame systems would eventually be integrated into a 3D frame system of "front-back, up-down, left-right", whose front-back direction is perpendicular to the canvas. [Figure 5.2.1-left]

The cell-scene image is a composition of the object image and the background image. According to the analysis in 4.2.1, parallelograms (or half parallelograms) appear as the structural relationship between object images or in the visual structure of the object image. Like the prototype in [Figure 5.1.1], these parallelograms would conflict with the 3D orthogonal background framework and thus triggers the generation of corresponding perceptual force and counterforce. [Figure 5.2.1-right] Since the parallelograms of all object images in the same cell-scene image have the same inclination, the perceptual force induced by these parallelograms of these object images should have the same direction. Accordingly, the directions of the counterforces are also the same.

The background image is composed of the sky image and ground image, or only the ground image. The sky image has an orthogonal visual structure that is parallel to the background framework. Therefore it does not trigger perceptual forces. While, the ground image has a visual structure of parallelogram, which would arise a set of perceptual and counter forces. According to the analysis in 4.2.1, its visual structure is determined by the visual structure of the object image it bears, which makes its parallelogram structure have the same inclination as the object images. Then, the perceptual force generated because of the parallelogram of the ground image would have the same direction as the perceptual forces generated by the parallelogram of the object images. Accordingly, their counterforces would be the same as well.

Therefore, when the viewer's focus point moves onto a certain object image or ground image, he would feel the perceptual force and counterforce generated by the parallelogram in this image. The force attempts to push the parallelogram onto the ground. The counterforce subjects the viewer's visual system and drive him to move towards the right-rear and look down to the left, so that the rectangle on the ground can be seen as this parallelogram.

5.2.2 Orientation and position of the diegetic viewpoint and visual structure of cell-scene image

It is clear that when the viewer looks at an image, there is a process in visual perception that finds the simplest structural configuration of the image and associates it with the current image. From

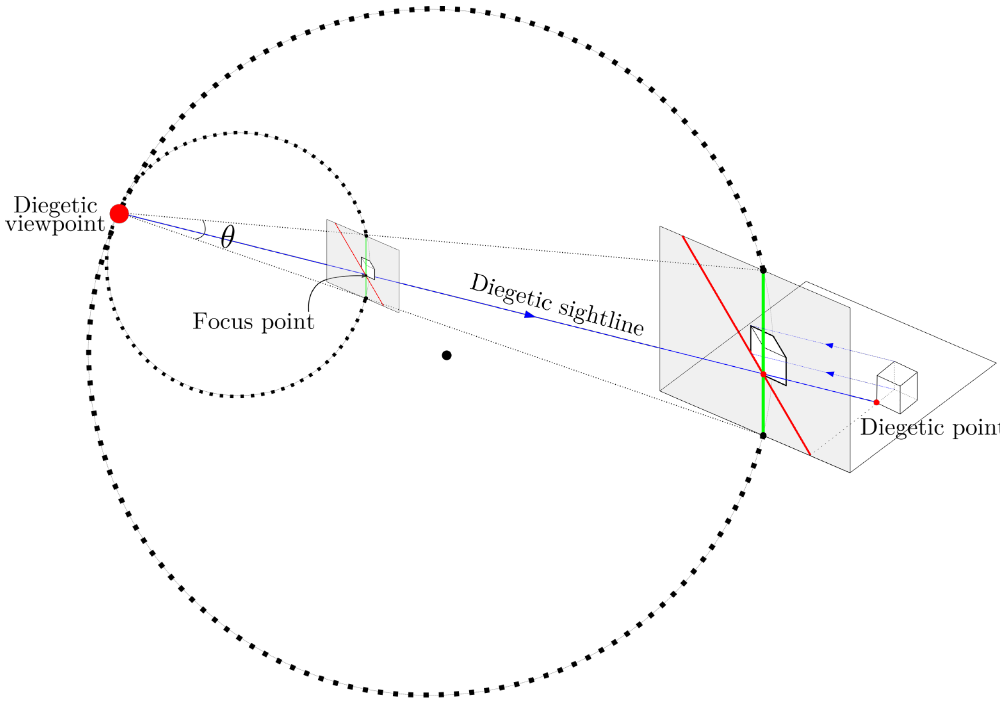

Figure 5.2.2: Position and orientation of the viewer's diegetic viewpoint

the previous analysis, this process is dependent on the visual experience of the viewer. Imagine if there is no second case in the viewer's visual experience, in which the rectangle in [Figure 5.1.1] can be seen as a parallelogram, then there is no possibility of associating the simplest structural configuration with the current image. The reason we can accept this association is that our visual experience tells us that there is such feasibility to associating the two together. This feasibility is the oblique projection, by which a rectangle on the ground can be projected as a parallelogram. And once we accept this association, what we are in fact accepting is the complete setup of this oblique projection. This means that we start to imagine placing ourselves in the position of the painter and keeping the same state as his. In this sense, the parallelogram establishes a connection between the viewer's and the painter's experiences.

According to the formation of the cell-scene image analyzed in 4.2.2, the parallelogram is also considered an oblique projection of a rectangle on the ground. That is to say, there exists a possibility in the visual experience of ancient Chinese painters to correlate a parallelogram with a rectangle on the ground. In ancient China, the painters and collectors of CHSP were the same groups of people. This means that the ancient CHSP viewers were familiar with projection techniques as well. Then, when they look at the parallelogram, they would also be able to correlate it with the rectangle on the ground. And when he accepted this setting, he in fact accepted the entire projection setup of the cell-scene image analyzed in 4.2.2. This includes the projection direction, the orientation of the viewpoint, the size reduction ratio, etc. This means that the viewer imagined placing himself in the situation where the painter was projecting the cell-scene image. [1] In this situation, the relationship between the viewer and the represented scene is equivalent to the relationship between the painter and the actual scene. Then, the

[1]This imaginary placement reflects the concept of representation in Chinese painting that, by setting up a specific visual structure, the painter directs the viewer to where he is standing and conveys his view to the viewer.

latter relationship can be illustrated as the diagram in [Figure 5.2.2] according to the diagram in [Figure 4.2.7].The blue line in the figure represents the viewer's diegetic sightline, which is on the same line of projection ray, but in the opposite direction. Based on the projection rule of the cell-scene image, the exact direction of the projection ray is related to the inclination of the diagonal passing through the projection point. Therefore, the direction of the diegetic viewpoint can be derived according to the inclination of the diagonal passing through the focus point.

According to the projection rule of the cell-scene image, the painter projected the cell-scene through a translucent projection plane (a piece of silk). Then, the painter's viewpoint was not only on the projection rays passing through the focus point but also on the intersection of any two projection rays passing through the two projection planes, as the intersection of two gray lines in [Figure 5.2.2].Then the position of the painter's viewpoint needs to be determined according to the intersection of the projection rays. Here I assume that the artist honestly projected the entire depth of the scene on this projection plane without cropping. In this situation, the small projection plane in [Figure 5.2.2] exactly obscures the vertical section of the painter's viewing frustum. As the sizes of the small and large projection planes follow the reduction principle of perspective, the height of the large projection plane is also exactly equal to the height of the vertical section of the painter's viewing frustum. If we abstract the painter's view as a viewing frustum with a constant Field of View. The position of the painter's viewpoint satisfying the assumed situation must be located on a special circle. This circle passes through the ends of the green vertical line crossing the focus point. And the inscribed angle θ corresponding to this green line is equal to the vertical angle of the painter's viewing frustum. Since the viewpoint is also located on the projection ray passing through the current focus point. Therefore, the painter's viewpoint is one of the intersections of the projection ray and the circle, which is located on the front side of the projection plane. Since the viewer's diegetic viewpoint coincides with the painter's, its exact position can be derived in this way.

5.3 Visual structure of column-scene image and movement of the diegetic viewpoint

The column-scene image is a vertical composition of multiple cell-scene images whose visual structures have diagonals with different inclinations. According to the geometric diagram in [Figure 4.2.7], the direction of the projection ray is directly related to the inclination of the diagonal passing through the focus point and each inclination corresponds to a projection direction.Therefore, a column-scene image corresponds to multiple projection directions. As the viewer's diegetic viewpoint has the opposite direction of the projection ray passing through his focus point on the painting, when the viewer's focus point moves vertically on the column-scene image, the direction of his diegetic viewpoint will undergo a change because the switch of cell-scene images. According to the analysis in 5.2.2, the viewer's diegetic viewpoint is located at the intersection of a special circle and the projection ray (the diegetic sightline) passing through his focus point. Then, when the position of the focus point and the direction of the projection ray (the diegetic sightline) change, their intersection (the position of the diegetic viewpoint) changes synchronously. [Figure 5.3.1]

The composition of the column-scene image determines the switching way of cell-scene images and further influences the movement of the diegetic viewpoint. In the development of the column-scene image, three compositions were developed. These three compositions result in three ways of switching between cell-scene images. These switches further determine the three ways of changing the direction and position of the viewer's diegetic viewpoint, when his focus point moves vertically on the column-scene image.

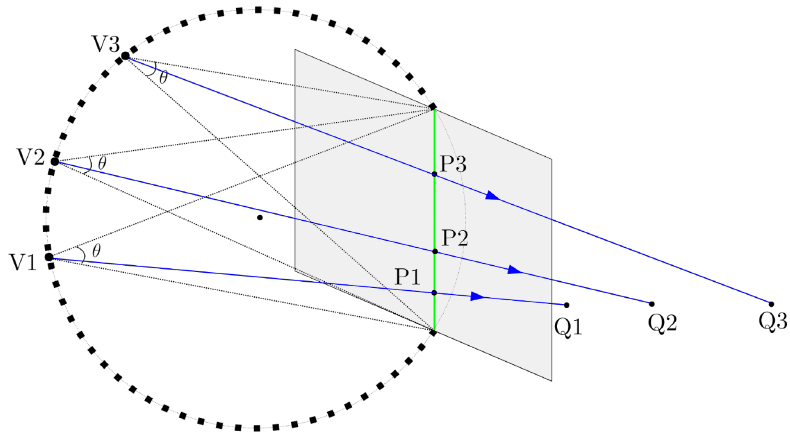

Figure 5.3.1: Visual structure of column-scene image and movement of the viewer's diegetic viewpoint

5.3.1 Diegetic movement corresponds to the first staged column-scene image

The composition of the column-scene image at the first stage is a compartmental juxtaposition of three cell-scene images. Usually, the visual structures of three cell-scene images have three parallelogram grids with different inclinations. They are stitched together without any structural transition. Therefore, when the viewer's focus point moves vertically on the column-scene image, the direction of the diegetic viewpoint would suddenly change at the boundaries, due to the sudden switches between the parallelogram grids of cell-scene images. As the viewer's diegetic viewpoint is located at the intersection of a special circle and the projection ray (the diegetic sightline) passing through his focus point, these changes in the direction of the diegetic viewpoint would cause a synchronous change in the position of the diegetic viewpoint.

Assuming that the vertical line AB in [Figure 5.3.2] is the moving path of the viewer's focus point across a column-scene image of the painting of *Goddess Luo Rhapsody*. Because of the compartmental composition, this path can be divided into three parts: the part in the upper cell-scene image (between A and B); the part across the boundary between the two images (at B); and the part in the lower cell-scene image (between B and C). When the viewer's focus point moves between AC, his diegetic viewpoint will undergo two types of movements. (Since the diagonals passing through the points between AB change in the same way as the diagonals passing through the points between BC, they correspond to the same movement.)

a. The movement of the diegetic viewpoint corresponds to AB (or BC).

The points between A and B are in the same cell-scene image, the diagonals passing through them then have the same inclination. Since the direction of the projection ray is related to the inclination of the diagonal, the projection rays passing through these points have the same direction. When the viewer's focus point moves between A and B, his diegetic viewpoint will move synchronously while keeping its direction unchanged. According to the inference in 5.2.2, the viewer's viewpoint is located at the intersection of the special circle and the diegetic sightline (projection ray). Then, as the viewer's focus point moves, the diegetic viewpoint slides on the arc from A' to B'. Since the points between the B and C are also located in the same cell-scene image, the viewer's diegetic viewpoint will conduct the same movement, as the focus point moves between B and C. [Figure 5.3.2]

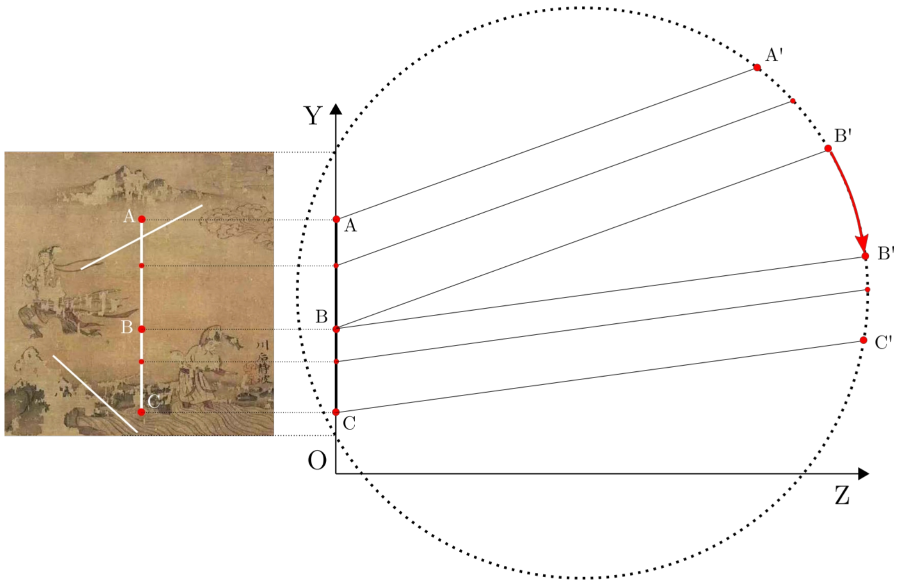

Figure 5.3.2: Movement of the viewer's diegetic viewpoint and visual structure of the column-scene image of the first stage

b. The movement of the diegetic viewpoint corresponds to B.

The points on either side of B are located in two cell-scene images whose diagonals have different inclinations. According to the composition of the column-scene at this stage, the visual structures of two adjacent cell-scene images are stitched together without any structural transition. Therefore, the diagonals passing through the points on either side of B suddenly rotate at B. This means that the directions of the projection rays passing through these points will also change at B. When the viewer's focus point passes through point B, the direction of the diegetic viewpoint will change suddenly from the direction that corresponds to A to that corresponds to B. Consequentially, the position of the diegetic viewpoint (the intersection of the special circle and the projection ray) will skip directly from the upper B' to the lower B'. [Figure 5.3.2]

5.3.2 Diegetic movement corresponds to the second staged column-scene image

The column-scene image in the second stage is generally a vertical composition of cell-scene images. However, there is an overlap between some adjacent cell-scene images. According to the analysis in 4.3.2, the overlapping region would be perceived to have a transitional visual structure. In this visual structure, the hidden diagonals gradually sway between the diagonals of the two adjacent cell scene images above and below. This means that the directions of the projection rays corresponding to these diagonals change gradually between the directions of the projection rays passing through the upper and lower diagonals. Then, when the viewer's focus moves vertically in this area, the direction of his diegetic viewpoint will also change gradually. Accordingly, the position of the diegetic viewpoint would synchronous changes, as it is located at the intersection of the special circle and the projection ray.

Assume that the line AD in [Figure 5.3.3] is the moving path of the viewer's focus point across the column-scene images of the painting of *Recumbent tour in Xiaoxiang*. According to the composition of this column-scene image, the moving path can be divided into three parts: the

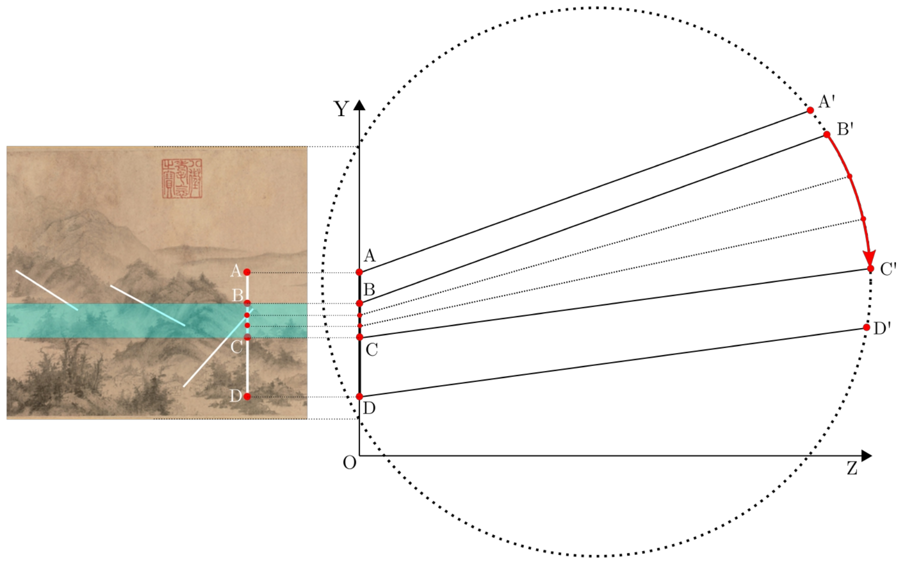

Figure 5.3.3: Movement of the viewer's diegetic viewpoint and visual structure of the column-scene image of the second stage

part in the upper cell-scene image (between A and B), the part in the overlapping area (between B and C), and the part in the lower cell-scene image (between C and D). When the viewer's focus point moves between AD, his diegetic viewpoint will undergo two types of movements. (Since the diagonals passing through the points between AB change in the same way as the diagonals passing through the points between CD, they correspond to the same movement.)

a. The movement of the diegetic viewpoint corresponds to AB (or CD).

The points on the path between A and B (or C and D) are in the same cell-scene image, then the directions of projection rays passing through these points are the same, as well as the directions of the projection rays (diegetic sight-lines). Therefore, when the viewer's focus point moves between A and C (or C and D), his diegetic viewpoint will move synchronously while keeping its direction unchanged. Consequentially, the position of the diegetic viewpoint (the intersection of the special circle and the projection ray) will slide on the arc from A' to B' (or from C' to D'). [Figure 5.3.3]

b. The movement of the diegetic viewpoint corresponds to BC.

The points between B and C are located in the overlapping area of two adjacent cell-scene images. As inferred, the inclinations of the diagonals passing these points vary gradually between the inclination corresponding to B and that corresponding to C. According to the diagram in [Figure 5.2.2], different inclinations means the different direction of diegetic sightlines (projection rays). When the viewer's focus point moves from B to C, the projection rays passing through these points will conduct a gradual rotation from the projection ray passing through B to the projection ray passing through C, due to the gradual change in the inclinations of diagonals passing the points from B to C.Consequentially, the position of the diegetic viewpoint (the intersection of the special circle and the projection ray) will slide synchronously on the arc from B' to C'. [Figure 5.3.3]

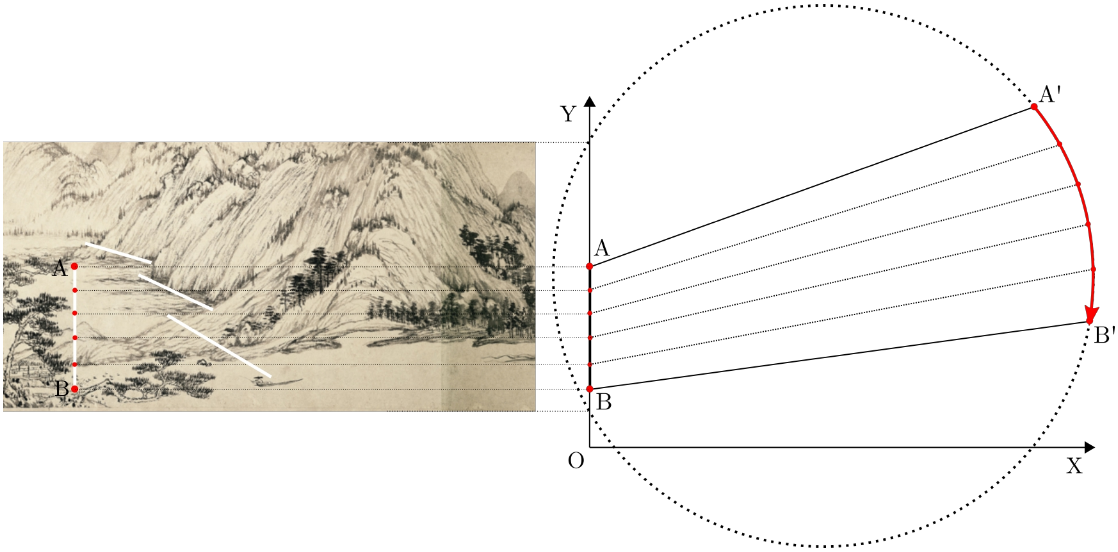

Figure 5.3.4: Movement of the viewer's diegetic viewpoint and visual structure of the column-scene image of the third stage

5.3.3 Diegetic movement corresponds to the third staged column-scene image

The column-scene image at this stage is characterized by a coherent object complex image, such as a mountain complex or riverbank, running vertically throughout the entire column-scene image. This complex object image usually has continuous size reduction or diagonal rotation, which implies that the object complex image is a combination of numerous single object images with the inclinations of their diagonals gradually changing. The entire column-scene image is therefore perceived as a continuous composition of numerous cell-scene images whose diagonals gradually rotate. This also means that the directions of the projected rays passing through these points change gradually. Therefore, when the viewer's focus point moves vertically on the column-scene image, the direction of his diegetic viewpoint will gradually change. Accordingly, the position of the diegetic viewpoint would synchronous changes, as it is located at the intersection of the special circle and the projection ray.

Assume that the line AB in [Figure 5.3.4] is the path of the viewer's focus point vertically moving among cell-scene images. The diagonals passing through A and B have different inclinations. According to the composition, the diagonals passing through the points between A and B rotate gradually between the diagonal passing through A and the diagonal passing through B. This means that the projection rays passing through these points also rotate gradually between the projection ray passing through A and the projection ray passing through B. Therefore, when the viewer's diegetic viewpoint moves along AB, the direction of his diegetic viewpoint will change gradually from the direction of the sightline (projection ray) passing through A to that of the sightline (projection ray) passing through B. Accordingly, the position of the diegetic viewpoint will synchronously slide on the arc from A' to B'.

In conclusion, the curves of these three stages of change show that the ancient painters intended to make the viewer's diegetic viewpoint move more and more uniformly when his focus point moved in a vertical direction across the CHSP.

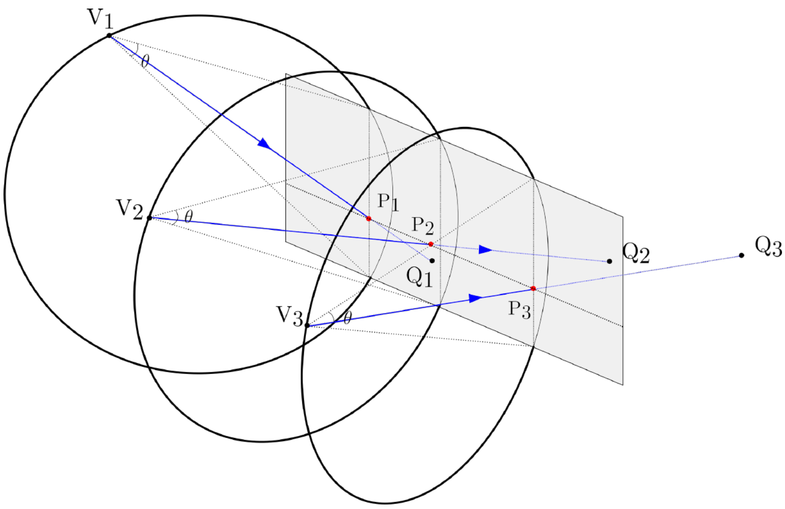

Figure 5.4.1: Visual structure of row-scene image and movement of the diegetic viewpoint

5.4 Visual structure of row-scene image and movement of the diegetic viewpoint

The row-scene image is a horizontal combination of multiple column-scene images. Each column-scene image is a vertical combination of cell-scene images. Therefore, the composition of the column-scene images is essentially the combination of cell-scene images in the horizontal direction. Usually, cell-scene images located at the same height have different visual structures, and the diagonals in these visual structures have different inclinations. According to the diagram in [Figure 5.2.2], the inclination of the diagonal corresponds to the projection direction. Therefore, the difference in diagonal inclinations means that the directions of the projection rays passing through these cell-scene images are also different.

The viewer's diegetic viewpoint has the opposite direction of the projection ray passing through his focus point on the painting. When the viewer's focus point moves horizontally across the row-scene image, the direction of his diegetic viewpoint will undergo a change because of the switch of cell-scene images. Accordingly, the position of the diegetic viewpoint would synchronous changes, as it is located at the intersection of the special circle and the diegetic sightline (projection ray). [Figure 5.4.1]

The composition of the row-scene image determines the switching method of cell-scene images in the horizontal direction, which would further affect the movement of the diegetic viewpoint. In the development of the row-scene image, three compositions were developed. These three compositions result in three ways of switching between column-scene images. These switches further determine the three ways of changing the direction and position of the viewer's diegetic viewpoint, when his focus point moves horizontally on the row-scene image.

5.4.1 Diegetic movement corresponds to the first staged row-scene image

The composition of the row-scene image at this stage is a compartmental juxtaposition of column-scene images. It is characterized by the presence of a distinct boundary between two adjacent

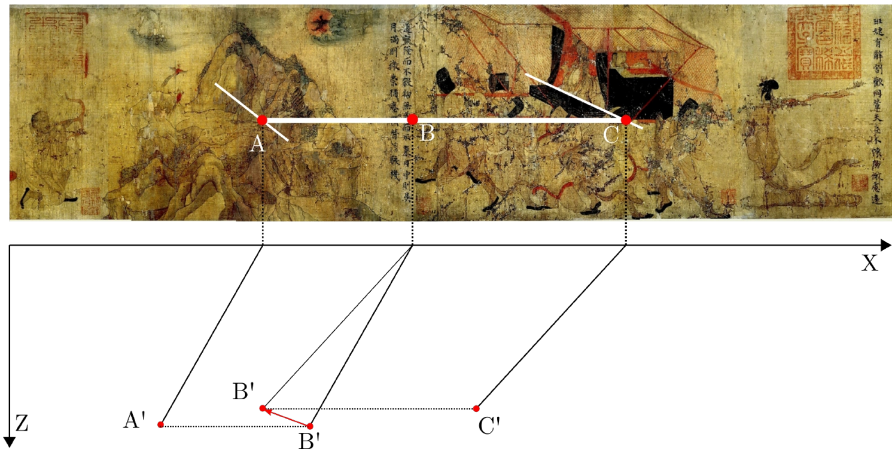

Figure 5.4.2: Movement of the diegetic viewpoint and visual structure of the row-scene image of the first stage

cell-scene images at the same height in the horizontal direction, such as vertically aligned text.[2] Their visual structures show direct stitching in the horizontal direction without any transition. In most cases, the two adjacent cell-scene images have different visual structures, in which the diagonals have different inclinations. This direct stitching then causes the diagonal lines of adjacent cell-scene images to rotate abruptly at the boundary. This abrupt rotation also implies an abrupt change in the direction of the projection rays passing through the cell-scene images on both sides. Thus, when the viewer's focus point moves horizontally across the row-scene image, the direction of his diegetic viewpoint will change suddenly at the boundary. Accordingly, the position of his diegetic viewpoint changes synchronously, as it is located at the intersection of a special circle and the diegetic sightline (projection ray) passing through the focus point.

Assume that the AB in [Figure 5.4.2] is the moving path of the viewer's focus point across a segmented row-scene image of the painting of *Admonitions of the Court Instructress.*According to its compartmental composition, the path can be divided into three parts: the part in the left column-scene image (between A and B), the part at the boundary between the two images (at B); and the part in the right column-scene image (between B and C). When the viewer's focus point moves between AC, his diegetic viewpoint will undergo two types of movements. (Since the diagonals passing through the points between AB change in the same way as the diagonals passing through the points between BC, they correspond to the same movement.)

a. The movement of the diegetic viewpoint corresponds to AB (or BC).

As the points between A and B are in the same cell-scene image, the diagonals passing through these points have the same inclinations. This means that the projection rays passing through these points have a unified direction. Thus, when the focus point moves between A and B (or B to C), the direction of his diegetic viewpoint will keep unchanged. However, as the diegetic viewpoint is located at the intersection of the special circle and the diegetic sightline (projection ray), the horizontal movement of the focus point will lead the diegetic viewpoint to move synchronously along a parallel line from A' to B' (or B' to C'). [Figure 5.4.2]

b. Movement of the diegetic viewpoint corresponds to B.

[2] The column-scene image is a vertical composition of cell-scene images

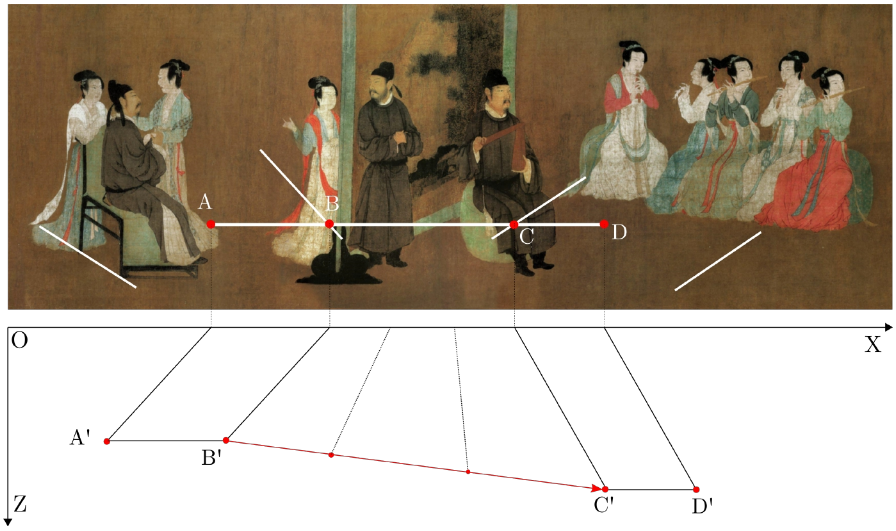

Figure 5.4.3: Movement of the viewer's diegetic viewpoint and visual structure of the row-scene image of the second stage

The points on either side of B are located in two cell-scene images whose diagonals have different inclinations. According to the composition of the row-scene image at this stage, the visual structures of two adjacent cell-scene images are stitched together without any structural transition. Therefore, the inclinations of the diagonals passing through the points on either side of B would suddenly change at B. This means that the directions of the projection rays passing through these points also suddenly change at B. When the viewer's focus point passes through point B, the direction of his diegetic viewpoint will change suddenly from the direction that corresponds to A to that corresponds to B. Consequentially, the position of the diegetic viewpoint (the intersection of the special circle and the projection ray) will skip directly from right B' to the left B', as shown in [Figure 5.4.2].

5.4.2 Diegetic movement corresponds to the second staged row-scene image

The row-scene image at this stage has a combined feature that a shared region that belongs to two adjacent column-scene images appears simultaneously, such as the area where two facing humans occupy in the [Figure 4.4.4], or the place on which a screen is located [Figure 4.4.5]. The feature implies a partial overlap of two adjacent column-scene images. As the column-scene image is a vertical combination of cell-scene images, this overlap is essentially the partial overlap of the cell-scene images at the same height as the row-scene image. According to the analysis in 4.4.2, the diagonals in the visual structure of the overlapping area would be perceived as gradually swinging between the diagonals of the two adjacent cell-scene images on the left and right. This means that the direction of the projection rays corresponding to these diagonals changes gradually between the directions of the projection rays passing through the left and right diagonals. Then, as the viewer's focus point moves horizontally across the row-scene image, the direction of his diegetic viewpoint changes gradually。Accordingly, the position of the diegetic viewpoint would synchronous changes, as it is located at the intersection of the special circle and the projection ray.

Assume that the line AD in [Figure 5.4.3] is the moving path of the viewer's focus point across a segmented row-scene image of the painting of *The Night Revels of Han Xizai.* According to the composition of this row-scene image, the moving path can be divided into three parts: the part in the upper cell-scene image (between A and B), the part in the overlapping area (between B and C), and the part in the lower cell-scene image (between C and D). When the viewer's focus point moves between AD, his diegetic viewpoint will undergo two types of movements. (Since the diagonals passing through the points between AB change in the same way as the diagonals passing through the points between CD, they correspond to the same movement.)

a. The movement of the diegetic viewpoint corresponds to AB (or CD).

The points on the path between A and B (or C and D) are in the same cell-scene image, then the directions of projection rays passing through these points are the same, as well as the directions of the projection rays (diegetic sightlines). Therefore, when the viewer's focus point moves between A and C (or C and D), his diegetic viewpoint will move synchronously while keeping its direction unchanged. Consequentially, the position of the diegetic viewpoint (the intersection of the special circle and the projection ray) will change synchronously. Its moving trajectory will be a line from A' to B' (or from C' to D') parallel to AB (or CD). [Figure 5.4.3]

b. The movement of the diegetic viewpoint corresponds to BC.

The points between B and C are located in the overlapping area of two adjacent cell-scene images. As inferred, the inclinations of the diagonals passing these points vary gradually between the inclination corresponding to B and that corresponding to C. These rotating diagonals further imply that the projection rays passing through these points between B and C rotate gradually.

When the viewer's focus point moves from B to C, the diegetic sightlines (projection rays) passing through these points will conduct a gradual rotation from the diegetic sightline passing through B to the diegetic sightline passing through C. Consequentially, the position of the diegetic viewpoint (the intersection of the special circle and the projection ray) will move synchronously from B' to C'. [Figure 5.4.3] As the direction of the diegetic sightline change gradually, the trajectory of the diegetic viewpoint will be a line between B' and C'.

5.4.3 Diegetic movement corresponds to the third staged row-scene image

The visual structure of the row-scene images at this stage is characterized by the swaying diagonals between two adjacent column-scene images in the horizontal direction. Due to the reification principle of Gestalt, these diagonals imply that there are numerous diagonals composed horizontally between two column-scene images with inclinations that gradually change with a tiny angle each time. According to the diagram in [Figure 5.2.2], different inclinations imply different directions of diegetic sightlines (projection rays). The directions of the projection rays passing through points at the same height gradually change, and the difference between any two adjacent ones is very small since the projection rays are numerous. Then, when the viewer's focus moves horizontally, his diegetic viewpoint will gradually rotate at a slight angle at a time. The position of his diegetic viewpoint will synchronously change, as it is located at the intersection of the special circle and the diegetic sight-line (projection ray). Ans its moving trajectory will show as a smooth curve because of the small angular difference between every two sightlines.

Assume that the line AB in [Figure 5.4.4] is a horizontal path of the viewer's focus point moving across a segmented row-scene image of *Spring morning in the Han Palace.* A and B are respectively in two cell-scene images at the same height. According to the composition, the diagonals passing through the points between A and B rotate gradually at a slight angle at a time

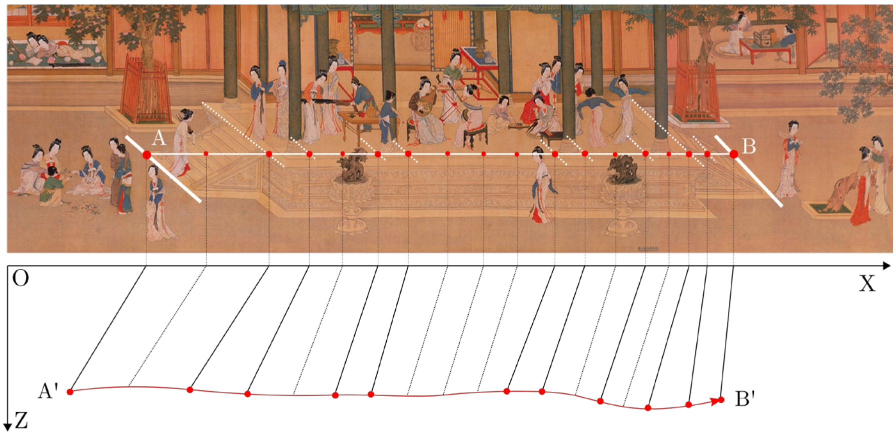

Figure 5.4.4: Movement of the viewer's diegetic viewpoint and visual structure of the row-scene image of the third stage

between the diagonal passing through A and the diagonal passing through B. This means that the projection rays passing through these points also rotate gradually at a slight angle at a time between the projection ray passing through A and the projection ray passing through B. Then, when the viewer's diegetic viewpoint moves along AB, the direction of his diegetic viewpoint will change gradually from the direction of the sightline (projection ray) passing through A to that of the sightline (projection ray) passing through B. Accordingly, the position of the diegetic viewpoint will synchronously change. As the direction of the diegetic sightline changes gradually between multiple angles, its trajectory will show as a smooth curve between A' and B'.

In conclusion, the curves of these three stages of change show that the ancient painters intended to make the viewer's diegetic viewpoint move more and more uniformly when his focus point moved in a horizontal direction across the CHSP.

Summary

This chapter analyzed the geometric correlation between the visual structure of the CHSP and diegetic movement.

The parallelograms in the CHSP visual structure provide the perceptual force for the viewer's movement in the diegetic world. Since object images and ground images contain parallelograms, a perceptual force can also be generated when viewing a cell-scene image. Also, because both the visual structure of the object image and the visual structure of the ground image have parallelograms with the same inclination, the perceptual forces triggered by these parallelograms are the same, as well as the diegetic movements they drive.

Visual structure of cell-scene image and direction and position of the diegetic viewpoint

The viewer's diegetic viewpoint locates on the projection ray that passes through the focus point and orients in the opposite direction as the projection. The position of the diegetic viewpoint is located at the intersection of the projection ray and a special circle on the front side of the projection plane. This circle passes through the two ends of the vertical line passing through the focus point on the projection plane. The inscribed angle corresponding to this vertical line equals the vertical angle of the painter's frustum.

Visual structure of column-scene image and movement of the viewer's diegetic viewpoint

The column-scene image is a vertical combination of cell-scene images with different visual structures. When the viewer's focus point moves in the vertical direction, the diegetic sightline (projection ray) where his diegetic viewpoint locates changes because of the switching of cell-scene images. The combination of cell-scene images determines the switching way between cell-scene images, which further affects the changing way of direction and position of the diegetic viewpoint.

The composition of the column-scene image at the first stage is a vertical compartmental juxtaposition of cell-scene images. When the focus point vertically moves in the same cell-scene image, the diegetic viewpoint moves on the same circle while keeping the direction of the sight-line unchanged, because the diagonals passing through the path points have the same inclination. When the viewer's focus point passes through the junction of two cell-scene images, the direction of the diegetic viewpoint abruptly changes, since the directions of the diagonals passing through the path points suddenly change. The diegetic viewpoint synchronously moves on the circle and the moving path suddenly becomes longer.

The composition of the column-scene image at the second stage is generally a vertical compartmental juxtaposition of cell-scene images, but some adjacent cell-scene images partly overlapped. When the focus point vertically moves in the same cell-scene image, the diegetic viewpoint moves on the same circle while keeping the direction of the sight-line unchanged. When the focus point vertically moves in the overlapping area, the direction of the diegetic viewpoint changes gradually, since the inclinations of the diagonals in the overlapping area gradually change between the inclinations in the upper and lower cell-scene images. The diegetic viewpoint synchronously moves on the circle, but the curved distance it moves each time does not evenly.

The composition of the column-scene image at the third stage is an open fusion of numerous cell-scene images. When the viewer's focus point moves along a vertical path through multiple cell-scene images in the column scene image, the direction of the viewer's diegetic viewpoint changes gradually and continuously with a tiny angle each time, because the diagonal lines in these cell-scene images change gradually. The diegetic viewpoint synchronously moves on a circle, but the curved distance it moves each time does not evenly.

Visual structure of row-scene image and movement of the viewer's diegetic viewpoint

The row-scene image is a horizontal composition of multiple column-scene images. And at the same height, it is a vertical combination of cell-scene images. When the viewer's focus point moves horizontally on the row-scene images, it is moving through multiple cell-scene images at the same height. Then, the switching way of the cell-scene images determines the changing way projection rays change, which further affects the changing way of direction and position of the diegetic viewpoint.

The composition of the row-scene image of the first stage is the horizontal compartmental juxtaposition of column-scene images multiple with boundaries. When the focus point horizontally moves in the same cell-scene image, the diegetic viewpoint keeps the same direction and moves parallel with the focus point. When the viewer's focus point passes the junction of two cell-scene images, the direction and moving distance of the diegetic viewpoint change abruptly due to the abrupt switch of different diagonals.

The composition of the row-scene image at the second stage is generally a horizontal compartmental juxtaposition of column-scene images, but some adjacent column-scene images partly overlapped. When the focus point moves horizontally in the same cell-scene image, the diegetic viewpoint keeps the same direction and moves parallel with the focus point. When the focus point moves horizontally in the overlapping area, the direction of the diegetic viewpoint changes gradually and the moving trajectory of the diegetic viewpoint is a curve.

The composition of the row-scene image at the third stage is the continuous open fusion of numerous column-scene images in the horizontal direction. When the viewer's focus point moves along a horizontal path through the column scene image, the direction of the viewer's diegetic viewpoint changes gradually and continuously with a tiny angle each time. The entire moving trajectory of the diegetic viewpoint is a smooth curve.

Chapter 6

Visual structure and diegetic scene

This chapter will discuss the geometric correlation between the visual structure of the CHSP and the spatial structure of the represented scene. First, the triggering effect of the primordial form in the visual structure of the CHSP in generating depth will be explained from the perspective of visual perception. Second, this chapter will elaborate on the inference of the geometric correlation between the visual structure of the cell-scene image and its spatial structure, as well as the estimation of depth in the cell-scene. Last, this chapter discusses the effect of the visual structures of the CHSP's column-scene image and row-scene image on their spatial structures and depth variations, based on examples from three development stages.

6.1 Parallelogram and spatial depth

According to the law of simplicity of Gestalt, the visual system prefers to believe in the situation where the image has the simplest visual structure when this image has multiple interpretations. Based on this theory, if a figure can be seen as another figure with a simpler visual structure in a 3D case, or the structural relationship of the figure and its background framework in a 3D case is simpler than the relationship between the current figure and its background frame in 2D image, the visual system will be more inclined to believe in the latter one. And at this moment, the 3D spatiality is generated from this image. From here, it can be seen that depth arises as a result of a synthesis of visual structure and the law of simplicity. This result is based on two premises. First, there needs to be a possibility in the visual experience of the viewer to interpret the current image as a thing in a 3D scene. Second, the visual structure of the thing in this 3D scene, or the structural relationship of the thing and its background framework needs to be simpler than those in the current image.

The parallelogram shown in [Figure 6.1.1] is the primordial form in the visual structure of the CHSP. This parallelogram can be interpreted as two kinds of figures in different situations. One is a parallelogram in the 2D space. The other is the projection of a rectangle on the ground in the 3D case. Since the rectangle with an orthogonal visual structure is apparently more concise than the parallelogram with a tilted visual structure, the perceptual system is more inclined to believe the 3D situation. As a result, 3D spatiality is then created. This process is easy to accept for modern people because in our visual experience this parallelogram can be obtained by following the oblique projection rules. Thus, when we look at it, we are able to associate this

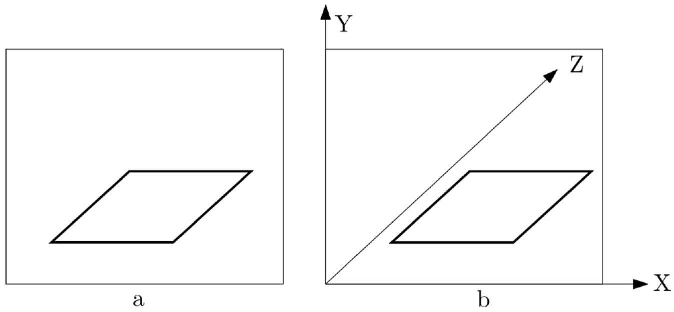

Figure 6.1.1: Parallelogram and spatial depth

parallelogram with a rectangle in the 3D scene based on our existing visual experience and are convinced of this association because of the law of simplicity. According to the ancient projection inferred in 4.2.2, the current parallelogram can also be interpreted as a projection of a rectangle on the ground. This implies that there was also a correlation between the parallelogram and the rectangle in the 3D scene in the visual experience of the ancient painters. Also, since the ancient viewers and painters were the same group, they shared the same visual experience. So when an ancient viewer would also naturally see this parallelogram as a rectangle on the ground in a 3D scene. It is by such correlation that ancient Chinese painters were able to represent three-dimensional space in two-dimensional paintings. They suggested depth by constructing parallelograms and built complex spatial scenes by combining them.

6.2 Spatial structure and spatial depth of cell-scene

6.2.1 Spatial structure of cell-scene

The cell-scene image is a combination of object images and the background image. Then, the spatial structure of the cell-scene is composed of the spatial structures of the object and background.

The background image of the cell-scene image has two types. One is the background image with only the ground image, as the background shows in [Figure 6.2.1-a]The other is the combination of the ground image and sky image. [Figure 6.2.1-c]According to the analysis in 4.2, the visual structure of the ground image in both two types is determined by the visual structure of the object image placed on it. They are perceived to have parallelogram structural grids. Because of the law of simplicity, these parallelograms will be interpreted as rectangular grids. Therefore, the spatial structure of the ground is a flat stage. The visual structure of the sky image is a rectangle. Its visual structure itself does not have the possibility to be interpreted as a simpler figure in a three-dimensional situation. Therefore it will still be perceived as a flat rectangle. However, since it connects to the highest part of the ground image, and this highest part corresponds to the farthest part in the 3D case, it will have a depth position in the 3D scene because of the three-dimensionality of the ground. By combining the ground and sky, the spatial structure of the cell-scene is perceived as a rectangular flat stage with a vertical background plane.

The object images also have two types of visual structure in the current CHSPs. One is the single flat form. The other is the composite form. The visual structure of the first type of object image itself can not be interpreted as a simpler structured figure in a 3D scene, so it will still be perceived as a planar form. However, they usually appear in groups, and their layout appears

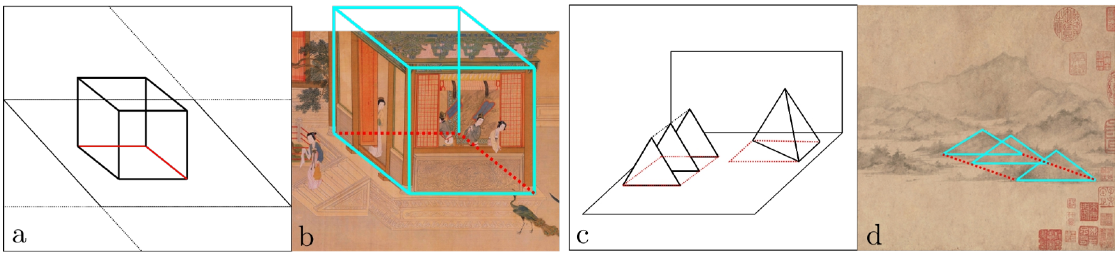

Figure 6.2.1: Parallelogram and spatial depth

like a parallelogram because of their staggering arrangement. For example, the mountain figures in the painting of *Recumbent Journey in Xiaoxiang*. [Figure 6.2.1-c]The parallelogram in their layout implies the distance in depth between them, thus making their combined spatial structure appear as multiple vertical billboards with different depths. The second type of object image usually has a parallelogram (or half parallelogram in some mountain images) in the base. This parallelogram implies its depth. The spatial structure of the object is then presented as the corresponding volume. Such as the building image in the painting of *Spring Morning in the Han Palace*, [Figure 6.2.1-b] Its visual structure will be perceived as a cube.

Combined with the above analyses, the spatial structure of the cell-scene is a flat stage with rectangular ground. The background of the stage is a vertical rectangle plane. In front of the background are placed vertical billboards or three-dimensional volumes with different depth distances.

6.2.2 Depth of cell-scene

According to the analysis in 6.1, the process of visual perception is a process of finding the most simplified structural configuration of an image and associating it with the current image. When the viewer accepts this simplest configuration, he is in fact accepting the entire setup of the projection of the cell-scene image analyzed in 4.2.2. That means that the viewer imagined placing himself in the situation where the painter was projecting the cell-scene image. In situation, the parameters of the viewer's diegetic viewpoint, such as the direction of the projection, the orientation of the viewpoint, and the size reduction scale, are the same as those of the painter's viewpoint. Thus, the relationship between the viewer and the represented scene is equivalent to the relationship between the painter and the actual scene. Then, the latter relationship can be illustrated as the diagram in [Figure 6.2.2] according to the same diagram in [Figure 4.2.7].

According to the inference in 4.2.2, the formation of the cell-scene image can be seen as a two-step process. The first step is to conduct a regular oblique projection. The second step is to scale down this projection according to the size of the object in the view. The distance between the diegetic viewpoint and any object in the cell-scene can be divided into three parts, according to these two operations. Of these, the blue part and the red part are the results of the two operations in the formation process of the cell-scene image. The blue part is the distance between a point in space and its projection point on the virtual projection plane. This distance corresponds to the oblique projection operation. The viewer's judgment of this distance follows the rule of oblique projection. According to this rule, this distance is related to two factors: the inclination of the diagonal passing through the projection point and the distance from this projection point to the ground along the direction of the diagonal (yellow line segment). [Figure 6.2.2]In the case of the constant yellow diagonal inclination, the greater its length, the greater this distance is. In the case of the constant length of the yellow line segment, the greater the

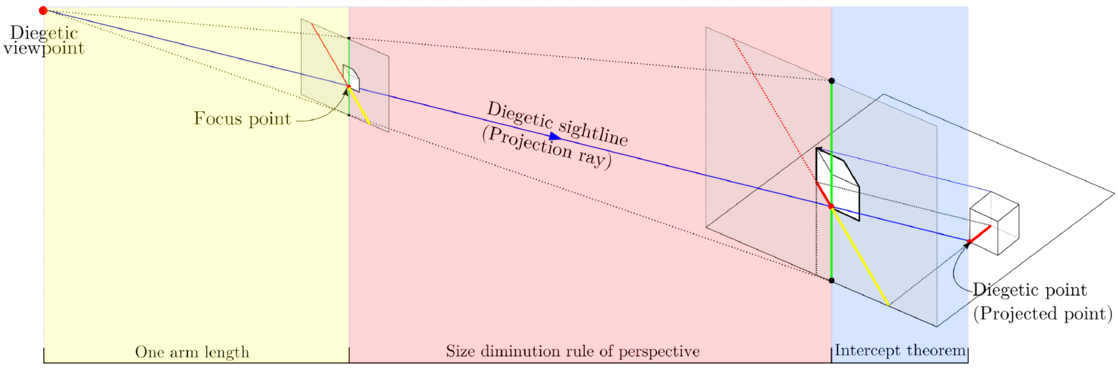

Figure 6.2.2: Estimation of spatial depth

angle between it and the ground, the greater this distance. The exact value of this distance can be derived from the length of the yellow line segment, the angle between it and the ground, and the angle between the projection ray and the virtual projection plane.[1]The red part is the distance between the projection point on the virtual projection plane and the point on the actual projection plane. This distance corresponds to the size reduction operation of perspective. The viewer's judgment of this distance is based on the ratio between the size of the object image in the painting (the size of the object image on the small projection plane) and the size of the actual object in his visual experience (the size of the object on the large projection plane). For example, the ratio between the lengths of the red diagonals on the two projection planes. According to the rule of perspective projection, when the angle of view (the angle between the two dashed lines in the figure) is constant, the larger the size reduction ratio between the sizes of the object images on the two projection planes, the greater the distance between them. The specific value of this distance can be calculated according to the size reduction ratio.[2] The yellow part is the distance between the viewpoint and the small projection plane. It is an arm‘s length in reality. As it is very small compared to the other two distances, this length can be ignored. Therefore, the distance between the diegetic viewpoint and the object is the red part plus the blue part.

6.3 Spatial structure and depth variation of column-scene

The column-scene image is a vertical composition of multiple cell-scene images. Since the spatial structure of a cell-scene is a flat stage with a rectangular ground, the spatial structure of a column scene is the combination of multiple flat stages in the depth. [Figure 6.3.1]In most cases, the sizes of the object images in the combined cell-scene images are reduced by different scales. This difference would be transferred to their respective ground images so that the unit length of the diagonal in the structural grid ground image represents a different actual length. According to the diagram in [Figure 6.2.2], the viewer's judgment of the depth of the cell-scene is based on the ratio[3] between the length of the diagonal in the painting (the length of the red diagonal on the small projection plane) and the length it represents in reality (the length of the red diagonal on the large projection plane).Since the column-scene image is a combination of multiple cell-scene images, there are multiple depth conversion ratios in the column-scene image. The way the cell-scene images are combined in the column-scene would affect the way the depth conversion ratios

[1]See the transformation of objection projection in Appendix B.
[2]See the transformation of perspective projection in Appendix B.
[3]This ratio is defined as the depth conversion ratio.

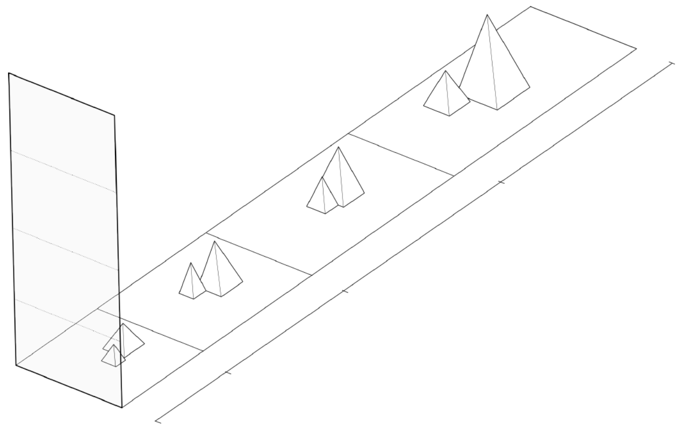

Figure 6.3.1: Spatial structure of column-scene

are switched. In the development of the column-scene image, three compositions were developed. These three compositions lead to three types of switching ways of depth conversion ratios, which ultimately result in three types of depth variations in the column-scene.

6.3.1 Spatial structure and depth variation of column-scene in the first stage

The column-scene image is the vertical juxtaposition of three cell-scene images. As the visual structure of each cell-scene image represents a cell-scene with a spatial structure of a flat stage, the spatial structure of the whole column-scene is a composition of three flat stages of the cell-scene along the depth direction. Typically, the visual structures of the three cell-scene images have different densities, and they are compartmentally juxtaposed without any structural transitions. This results in an abrupt change in the depth conversion ratio of the two adjacent cell-scene images at the boundary.

Assume that A, B, and C in [Figure 6.3.2] are points on a vertical line across a column-scene image of the painting of *Goddess Luo Rhapsody*, of which the segment AB and BC are in the upper and lower cell-scene images, respectively, and B is at the junction of the two cell-scene images. According to the diagram in [Figure 6.3.2], the depth conversion ratios corresponding to all points in a cell-scene image are the same. Then, the factor that determines the difference in depth between the 3D positions corresponding to these points is the different lengths of the diagonals starting from these points to the lower boundary of the cell-scene image. And since the lengths of the diagonals passing through these points are proportional to their heights, the depths corresponding to the points between A and B, or those between B and C, will vary linearly, as their heights increase. The depth variation curves corresponding to these points are then presented as two straight lines, as shown in [Figure 6.3.2-right].However, the degrees of size reduction of the mountain image and the figure image shows that the size reduction ratio in the cell-scene image above is larger than that in the one below. That means that the depth corresponding to the unit length of the diagonal (the depth conversion ratio) in the upper cell-scene image is greater than the depth corresponding to the unit length of the diagonal in the lower cell-scene image. This is reflected in the diagram that the slope of the depth variation line $A'B'$ corresponds to the points between A and B is greater than the slope of the depth variation line $B'C'$ corresponds to the points between B and C. Since the upper and lower cell-scene

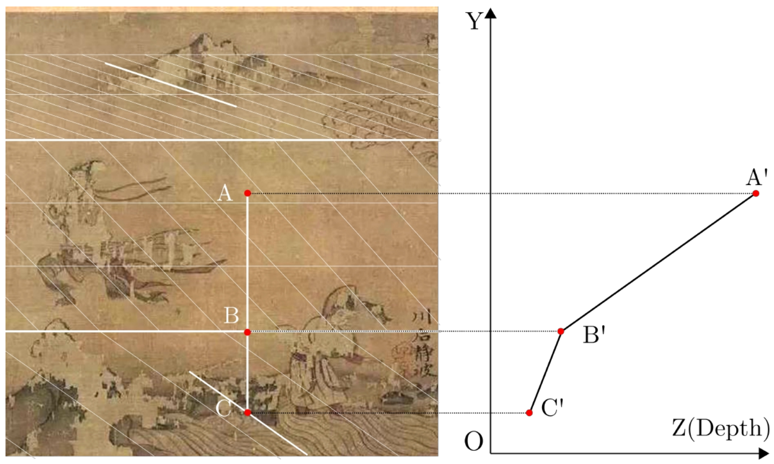

Figure 6.3.2: Depth variation in the column-scene image of the first stage

images are directly stitched without transition, the depth conversion ratios in the two images will switch abruptly at the boundary. This reflects the two depth variation lines would intersect at B'. [Figure 6.3.2-right]

6.3.2 Spatial structure and depth variation of column-scene in the second stage

The column-scene image in the second stage is generally a vertical composition of cell-scene images. However, partial overlap appears between some adjacent cell-scene images. Correspondingly, the spatial structure of the whole column-scene generally is a composition of multiple flat stages of cell-scene along the depth direction. However, there is a partial overlap in the depth direction between some neighboring cell-scenes. According to the analysis in 4.2.2, the visual structure of the overlapping area would be perceived as a transitional visual structure between the upper and lower cell-scene images. One of the features of this transitional visual structure is that the size reduction ratio corresponding to the points in this overlapping area varies gradually between the size reduction rations of the adjacent cell-scene images. This variation means that the depth represented by the unit length of the diagonals (depth conversion ratio) passing through these points varies gradually between the two.

Assume that line AD in [Figure 6.3.3] is a vertical line across two adjacent cell-scene images in a column-scene image of the painting of *Recumbent Tour in Xiaoxiang*. Segments AB and CD are in the upper and lower cell-scene images, respectively. The points between A and B, or between C and D, correspond to a constant depth conversion ratio, as they are in the same cell-scene image. These points correspond to the same type of depth variation. Segment BC is in the overlapped area. The depth conversion ratios corresponding to the points between B and C vary between two constant depth conversion ratios in upper and lower cell-scene images. They correspond to another way of depth variation.

a. Depth variation in the area between A and B (or C and D)

As the depth conversion ratios corresponding to all points in a cell-scene image are the same, the factor that determines the difference in depth between the 3D positions corresponding to these points is the different lengths of the diagonals starting from these points to the lower

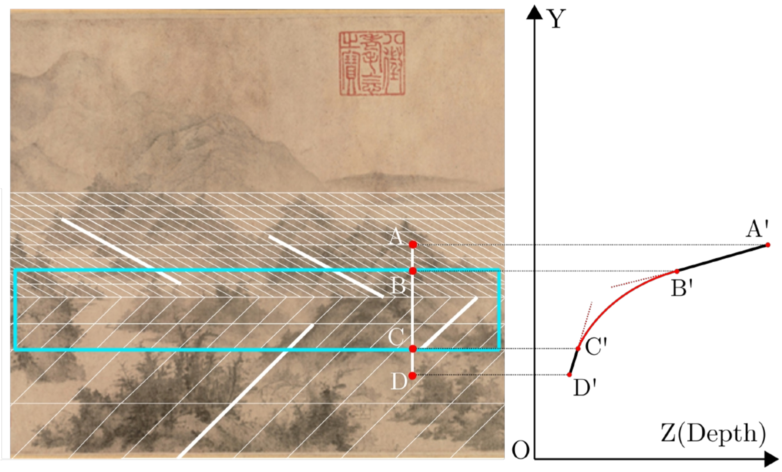

Figure 6.3.3: Depth variation in the column-scene image of the second stage

boundary of the cell-scene image. Since the inclinations of the diagonals passing through the points between A and B are the same, the lengths of these diagonals that start from these points to the lower boundary of the image are the same. Therefore, the depths that the points between A and B, or C and D, represent undergo a linear variation with their heights increase. Therefore, the depth variation curves corresponding to these two sets of points would present as two straight lines. However, the degrees of size reduction of the mountain images in the two cell-scene images show that the sizes of the mountain images in the upper one are more reduced than that in the lower cell-scene image. This indicates that the depth per unit length of the diagonal represents in the upper cell-scene image is larger than the depth per unit height in the lower cell-scene image. This is reflected in the diagram that, the slope of the depth variation line $A'B'$ corresponds to the points between A and B is greater than the slope of the depth variation line $C'D'$ corresponds to the points between C and D. [Figure 6.3.3]

b. Depth variation in the area between B and C

B and C are located in the upper and lower cell-scene images respectively, and the size reduction ratio in the upper cell-scene image is larger than that in the lower one. Therefore, the depth corresponding to the unit length of the diagonal (the depth conversion ratio) passing through B in the upper cell-scene image is greater than the depth corresponding to the unit length of the diagonal passing through C in the lower cell-scene image. This means that there is an increase in the depth conversion ratio corresponding to the points between B and C. [Figure 6.3.3]

As BC is located in the overlapping area, the depth conversion ratio corresponding to the points in this region varies gradually from that corresponding to B to that corresponding to C. Then, the increasing process of depth conversion ratios corresponding to the points between B and C is also gradual. According to the coordinations system of the diagram in [Figure 6.3.3], the depth conversion ratio of a point in the column-scene image is embodied as the slope of the tangent passing through its corresponding point on the depth variation curve.Then, the gradual change of the depth conversion ratios corresponding to the points between B and C is reflected as the gradual increase of the slopes of the tangents passing through their corresponding points on the depth variation curve. The points on the depth variation curve corresponding to the points between B and C form a smooth curve.

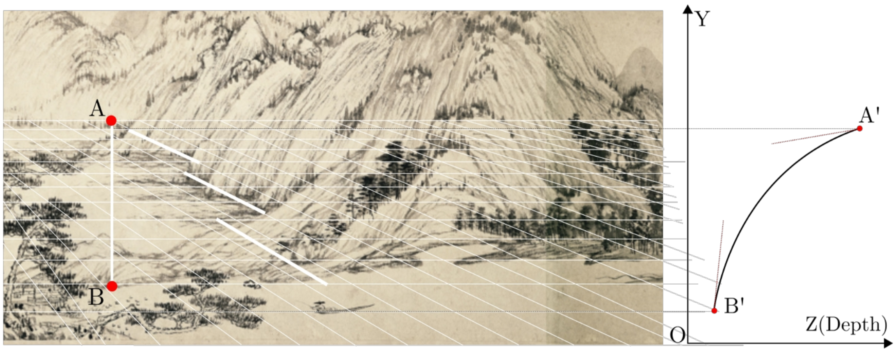

Figure 6.3.4: Depth variation in the column-scene image of the third stage

6.3.3 Spatial structure and depth variation of column-scene in the third stage

The column-scene image at this stage is characterized by a coherent object complex image, such as a mountain complex or riverbank, running vertically throughout the entire column-scene image. This complex object image usually has continuous size reduction or diagonal rotation, which implies that the object complex image is a combination of numerous single object images with their depth conversion ratios gradually changing. The entire column-scene image is therefore perceived as a continuous composition of numerous cell-scene images whose diagonals gradually rotate. Accordingly, the spatial structure of the column-scene image is a combination of numerous flat stages of cell-scene along the depth direction. According to the analysis of the visual structure of the column scene image at this stage in 4.3.3, the sizes of the combined single object images in the object image complex are gradually reduced. This visual structural property is also transferred to the global ground image due to the relationship between the ground image and the object image, which results in the ground image's structural grid having the same size reduction ratio. Therefore, the size reduction ratio and the depth conversion ratio corresponding to the points on the whole column-scene image will vary gradually.

Assume that A and B are two endpoints of a vertical line running through a column-scene image of the painting of *Dwelling in the Fuchun Mountains*. [Figure 6.3.4] According to the above analysis, AB goes through numerous cell-scene images. From the difference in the size reductions of the farthest and the nearest mountain image, it can be seen that the cell-scene images in which they are located use different size reduction ratios, and this ratio in the upper cell-scene image is larger than that in the lower one. That means that the depth corresponding to the unit length of the diagonal (the depth conversion ratio) in the upper cell-scene image is greater than the depth corresponding to the unit length of the diagonal in the lower cell-scene image. This is reflected in the diagram that the inclination of the depth variation line corresponding to the points between A and B is greater than the inclination of the depth variation line corresponding to the points between A and B. As the depth conversion ratio corresponding to the points in the whole column-scene image varies gradually from that corresponding to A to that corresponding to B. Then, the increasing process of the depth conversion ratio corresponding to the points on AB is gradual.

As the depth conversion ratio of a point in the column-scene image is embodied as the slope of the tangent passing through its corresponding point on the depth variation curve, the gradual change of the depth conversion ratios corresponding to the points between A and B is reflected as the gradual increase of the slopes of the tangents passing through their corresponding points

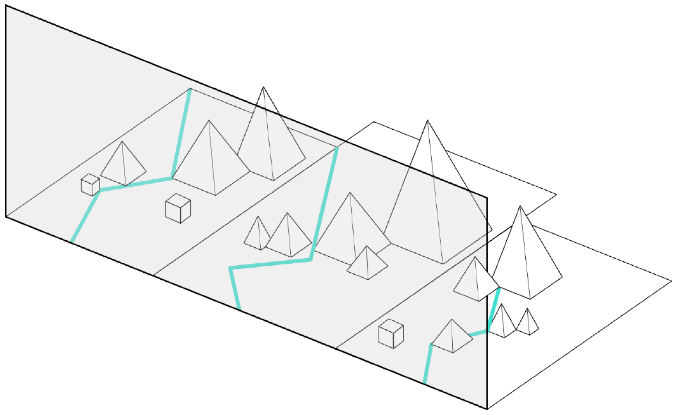

Figure 6.4.1: Spatial structure of row-scene

on the depth variation curve.The points on the depth variation curve corresponding to the points between A and B form a smooth curve. [Figure 6.3.4]

6.4 Spatial structure and depth variation of row-scene

The row-scene image consists of a horizontal combination of the multiple column-scene images. Each column-scene image is a vertical combination of cell-scene images. Therefore, the spatial structure of the column-scene is the combination of flat stages of cell-scenes in the depth direction and horizontal directions. [Figure 6.4.1]

In some row-scene images, the size reduction ratios of the object images in the cell-scene images located at the same height are different. This difference will be transferred to the ground image of the cell-scene images they are each in so that the diagonals per unit length in the ground image's visual grids represent different actual lengths. Or rather, their deep conversion ratios are different. Since the row scene image is essentially a combination of multiple cell-scene images in the horizontal direction, there are multiple depth conversion ratios at the same height of the row-scene image. The way the cell-scene images are combined at the same height as the row-scene would affect the way the depth conversion ratios are switched. In the development of the row-scene image, three compositions were developed. These three compositions lead to three types of switching ways of depth conversion ratios, which ultimately result in three types of depth variations in the row-scene.

6.4.1 Spatial structure and depth variation of row-scene in the first stage

The composition of the row-scene image at this stage is a compartmental juxtaposition of column-scene images. It is characterized by the presence of a distinct boundary between two adjacent cell-scene images[4] at the same height, such as vertically aligned text. Their visual structures are compartmentally juxtaposed without any transition. Therefore, the spatial structure of the row-scene in the horizontal direction is the juxtaposition of several flat stages of cell-scenes with different depths. In most cases, the two adjacent cell-scene images present different size

[4]The column-scene image is a vertical composition of cell-scene images

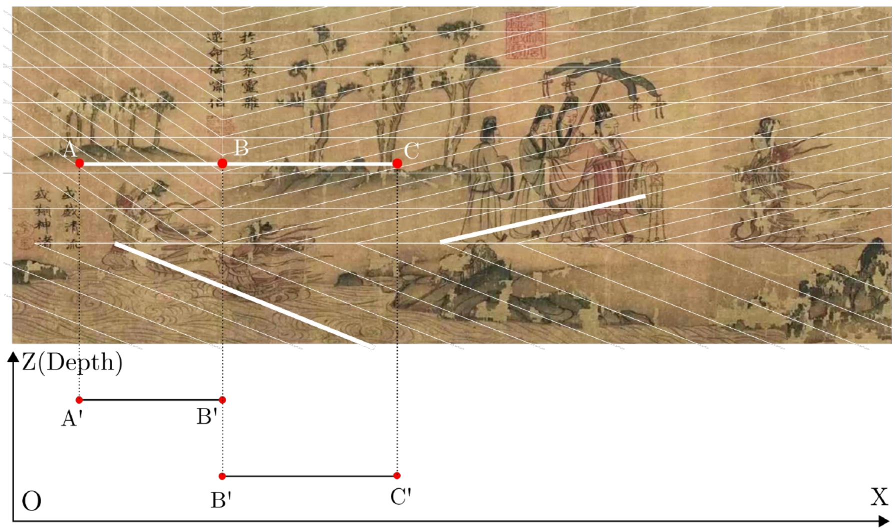

Figure 6.4.2: Depth variation in the row-scene image of the first stage

reductions, which implies they have different depth conversion ratios. Then, this compartmental juxtaposition would cause an abrupt shift of the size reduction ratios and depth conversion ratios at the boundary. This shift then implies an abrupt change in depths corresponding to points at the same height of the row-scene image at the boundary.

Assume that the line AC is a horizontal line segment across a segmented row-scene image of the painting of *Goddess Luo Rhapsody*. The segments AB and BC are in two adjacent cell-scene images, respectively, and B is at their boundary. Since all points in a cell-image correspond to the same depth conversion ratio, the factor that determines the difference in depth between the 3D positions corresponding to these points is the different lengths of the diagonals starting from these points to the lower boundary of the cell-scene image. Also, because the inclinations of the diagonals passing through the points between A and B are the same, the lengths of these diagonals that start from these points to the lower boundary of the image are the same. Therefore, the depths corresponding to the points between A and B are the same. As reflected in the diagram in [Figure 6.4.2], the depth variation curve corresponding to these points is a horizontal line. By the same token, the depth variation curve corresponding to the points between B and C is also a horizontal line.

The size difference between the tree images in the two adjacent cell-scene images shows that the size reduction ratio in the left cell-scene image is larger than that in the right one. That means that the depth corresponding to the unit length of the diagonal (the depth conversion ratio) in the left cell-scene image is greater than the depth corresponding to the unit length of the diagonal in the right cell-scene image. This is reflected in the diagram that the slope of the depth variation line $A'B'$ corresponds to the points between A and B is greater than the slope of the depth variation line $B'C'$ corresponds to the points between B and C. Since the left and right cell-scene images are directly stitched without transition, the depth conversion ratios in the two images will switch abruptly at the boundary. This reflects the two depth variation lines would intersect at B'. [Figure 6.4.2]

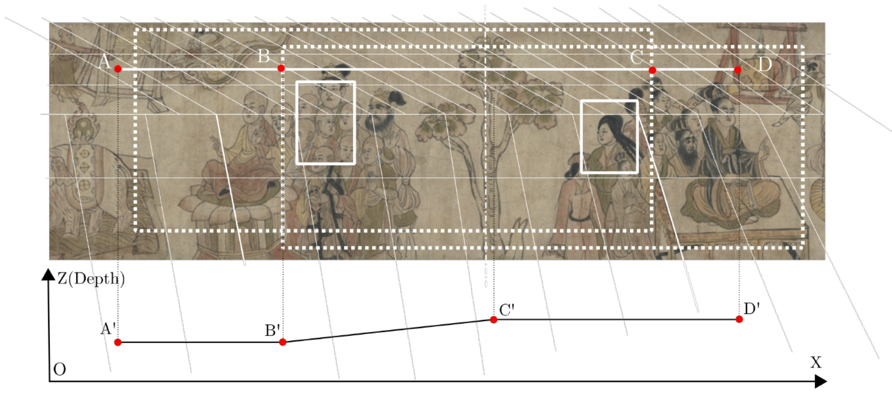

Figure 6.4.3: Depth variation in the row-scene image of the second stage

6.4.2 Spatial structure and depth variation of row-scene in the second stage

The row-scene image at this stage has a combined feature that a shared region that belongs to two adjacent column-scene images appears simultaneously, such as the area where two facing humans occupy in the [Figure 4.4.4], or the place on which a screen is located [Figure 4.4.5]. The feature implies a partial overlap of two adjacent column-scene images. As the column-scene image is a vertical combination of cell-scene images, this overlap is essentially the partial overlap of the cell-scene images that are at the same height as the row-scene image. Therefore, the spatial structure of the row-scene in the horizontal direction is the juxtaposition of several flat stages of cell-scenes. And some adjacent cell-scenes are partially overlapped. According to the analysis in 4.2.2, the visual structure of the overlapping area would be perceived as a transitional visual structure between the left and right cell-scene images. One of the features of this transitional visual structure is that the size reduction ratio corresponding to the points in this overlapping area varies gradually between the size reduction ratios of the adjacent cell-scene images. This variation means that the depth represented by the unit length of the diagonals (depth conversion ratio) passing through these points varies gradually between the two.

Assume that A, B, C, and D are points on a horizontal line across two adjacent column-scene images in a row-scene image of the painting of *Transformation on Vanquishing Mara*. [Figure 6.4.3]Both two adjacent column-scene images are vertically combined by two cell-scene images. Segment AB and CD are respectively in the left and right cell-scene images. The points on AB, or CD, correspond to a constant depth conversion ratio, as they are in the same cell-scene image. These points correspond to the same type of depth variation. Segment BC is in the overlapped area. The depth conversion ratios corresponding to the points between B and C vary between two constant depth conversion ratios in upper and lower cell-scene images. They correspond to another way of depth variation.

a. Depth variation corresponding to the points between A and B (or C and D)

As the depth conversion ratios corresponding to all points in a cell-scene image are the same, the factor that determines the difference in depth between the 3D positions corresponding to these points is the different lengths of the diagonals starting from these points to the lower boundary of the cell-scene image. Also, because the inclinations of the diagonals passing through the points between A and B are the same, the lengths of these diagonals that start from these points to the lower boundary of the image are the same. Therefore, the depths corresponding to the points between A and B are the same. As reflected in the diagram in [Figure 6.4.3], the depth variation curve corresponding to these points is a horizontal line. By the same token, the depth variation curve corresponding to the points between C and D is also a horizontal line.

The size difference between the drummer images in the two adjacent cell-scene images shows that the size reduction ratio in the left cell-scene image is larger than that in the right one.That means that the depth corresponding to the unit length of the diagonal (the depth conversion ratio) in the left cell-scene image is greater than the depth corresponding to the unit length of the diagonal in the right cell-scene image. This is reflected in the diagram that the slope of the depth variation line $A'B'$ corresponds to the points between A and B is greater than the slope of the depth variation line $C'D'$ corresponds to the points between C and D. [Figure 6.4.3]

b. Depth variation corresponding to the points between B and C

B and C are located in the left and right cell-scene images respectively, and the size reduction ratio in the left cell-scene image is larger than that in the right one. Therefore, the depth corresponding to the unit length of the diagonal (the depth conversion ratio) passing through B in the left cell-scene image is greater than the depth corresponding to the unit length of the diagonal passing through C in the right cell-scene image. This means that there is an increase in the depth conversion ratio corresponding to the points between B and C.

As BC is located in the overlapping area, the depth conversion ratio corresponding to the points in this region varies gradually from that corresponding to B to that corresponding to C. Then, the increasing process of depth conversion ratios corresponding to the points between B and C is also gradual. This is reflected that the points on the depth variation curve corresponding to the points between B and C would form a straight line connecting B' and C'. [Figure 6.4.3]

6.4.3 Spatial structure and depth variation of row-scene in the third stage

The row-scene image at this stage is characterized by the appearance of diagonal lines that gradually sway in the horizontal direction between the two adjacent column-scene images. This swaying phenomenon implies that there are numerous diagonals between these two column-scene images. Since each column-scene image is a vertical combination of multiple cell-scene images, this phenomenon also implies that there are numerous cell-scene images combined between these two cell-scene images at the same height. Therefore, the spatial structure of the row-scene in the horizontal direction is the juxtaposition of numerous flat stages of cell-scenes.

In some row-scene images, the size reduction ratios of the object images in the cell-scene images located at the same height are gradually reduced. This gradual reduction in size would be transferred to the ground image of the cell-scene images they are each in so that the actual lengths that the diagonals per unit length in the ground image's visual grids represent gradually change. Or rather, their deep conversion ratios gradually change. Therefore, the size reduction ratio and the depth conversion ratio corresponding to the points on the same horizontal line would vary gradually.

Assume that A and B are two endpoints of a vertical line running through a segmented row-scene image of the painting of *Shang Lin*. [Figure 6.4.4] According to the above analysis, AB goes through numerous cell-scene images. From the difference between the sizes of the building images near A and B, it can be seen that the cell-scene images where these build images are located were applied with different size reduction ratios, and this ratio in the left cell-scene image is larger than that in the right one. That means that the depth corresponding to the unit length of diagonal passing through A (the depth conversion ratio corresponding to A) in the left cell-scene image is greater than the depth corresponding to the unit length of the diagonal passing

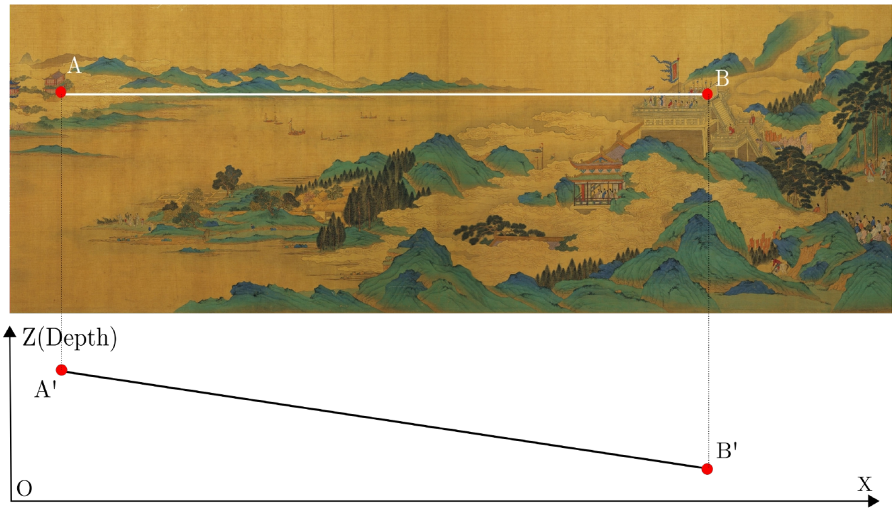

Figure 6.4.4: Depth variation in the row-scene image of the third stage

through A in the right cell-scene image. Since the heights of A and B are approximately the same, the length of the diagonal from them to the lower edge of the cell-scene image is roughly the same. Therefore, the depth corresponding to A is greater than the depth corresponding to B.[5]

Meanwhile, as the depth conversion ratio corresponding to the points in the whole row-scene image varies gradually from that corresponding to A to that corresponding to B. Then, the increasing process of depth conversion ratios corresponding to the points between A and B is also gradual. This is reflected that the points on the depth variation curve corresponding to the points between A and B would form a slanting line connecting A' and B'. [Figure 6.4.4]

Summary

This chapter established the geometric correlation between the visual structure of the CHSP and the spatial structure of the represented scene.

In classical Chinese paintings, the parallelogram in the CHSP visual structure suggested depth in the CHSP, because it could be seen as a rectangle on the ground. The ancient Chinese painter used this correlation to represent various spaces by composing different parallelograms. In the CHSP, parallelograms were composed into three lays of the CHSP, which correspondingly represent three kinds of spatial structures.

Spatial structure and depth of cell-scene

The spatial structure of the cell-scene is a flat stage with a rectangular base and a vertical rectangle background, on which receding 2D planar forms or 3D volumes are placed. The distance between the viewpoint and the object can be divided into three parts. The first part corresponds to the oblique projection operation in the formation of the cell-scene image. This part of distance

[5] According to the diagram in [Figure 6.2.2], the difference in the inclination of the diagonals passing through A and B also affects the judgment of the depths corresponding to these two points. But compared to the difference caused by the depth conversion ratio, the difference caused by this factor is small.

relates to the diagonal length and inclination, which can be calculated according to the intercept theorem. The second part corresponds to the size reduction operation. This part of distance relates to the ratio (size reduction ratio) between the size of the object image in the painting and its size in reality, which can be calculated according to the size diminution rule of perspective. The third part is the distance between the viewpoint and the small projection plane. It is an arm's length in reality. Compared to the other two distances, this length can be ignored.

Spatial structure and depth variation of column-scene

The spatial structure of the column-scene at the first stage is a combination of the flat stage in the depth direction. The depths corresponding to the points on a vertical line in the column-scene change linearly. At the junction of two adjacent cell-scene images, the depth variation speed change abruptly.

The spatial structure of the column-scene at the second stage is generally a composition of flat stages in the depth, but some neighboring cell-scenes are partially fused. The depths corresponding to the points on a vertical line in the overlapping area change non-linear, and normally the speed increases fast as the points' heights increase.

The spatial structure of the column-scene at the third stage is a continuous composition of numerous flat stages in the depth. The depths corresponding to the points on a vertical line in the entire column-scene image change non-linear, and normally the speed increases fast as the points' heights increase.

Spatial structure and depth variation of row-scene

The spatial structure of the row-scene at the first stage is a combination of the flat stages in the horizontal direction. The depths corresponding to the points on a horizontal line in the row-scene change linearly. At the junction of two adjacent column-scene images, the depth variation change abruptly.

The spatial structure of the row-scene is generally a horizontal combination of flat stages in the horizontal direction, but some adjacent column-scenes are partially fused. The depths corresponding to the points on a vertical line in the overlapping area change gradually.

The spatial structure of the row-scene at the third stage is a continuous composition of numerous flat stages in the horizontal direction. The depths corresponding to the points on a horizontal line in the entire row-scene image change gradually.

Part III

Computation & Construction

Chapter 7

Algorithm development for simulating diegetic experience

This chapter will elaborate on the algorithm for simulating the diegetic experience of "touring in the CHSP". First, the design of the algorithm with projection rays as module links will be explained. Second, three modules of the algorithm will be illustrated in detail: Module 1 for calculating the directions of the projection rays (diegetic sightlines) passing through all points on the CHSP, Module 2 for calculating the orientations and positions of the diegetic viewpoints corresponding to all points on the CHSP, Module 3 for constructing the diegetic scene corresponding to the image in the unrolled frame.

7.1 Algorithm design for simulating diegetic experience

The simulation of diegetic experience in fact consists of two tasks: directing diegetic movement and constructing the diegetic scene. To accomplish these two tasks, two sets of data are required: the orientation and position corresponding to all the points of the CHSP and the 3D positions of all the points of the CHSP. According to the diagrams in [4.2.2][5.2.2][6.2.2], both data are related to the projection ray (diegetic sightline)[1].First, the 3D positions of all points of the CHSP are located on the projection rays passing through them. Second, the diegetic viewpoint is oriented in the opposite direction to the projection ray and lies at the intersection of the projection ray and a special circle. Therefore, calculating the projection rays passing through all points on the CHSP is the key to obtaining the above data.

For this reason, I designed an algorithm for simulating the diegetic experience based on the projection ray. This algorithm is composed of three modules, as shown in [Figure 7.1.1].

Module 1 is to calculate the projection ray passing through all the points of the CHSP. The input of this module is the entire CHSP and depth clues. The calculation is based on the geometric relationship between the direction of the projection ray and the diagonal in the visual structure of the CHSP shown in the diagram of [Figure 4.2.7].The output is the data-table of the unit vectors of projection rays passing through all the points.

Module 2 is to calculate the position and orientation of the viewer's viewpoint corresponding to all the points of the CHSP. The input of this module is the data-table of unit vectors output from **Module 1** and the coordinates of the points of the CHSP. The calculation is based on the geometric correlation between the projection ray and the special circle shown in the diagram

[1]Projection rays and diegetic sightline have the same properties according to the analysis in 5.2.2 and 6.2.2

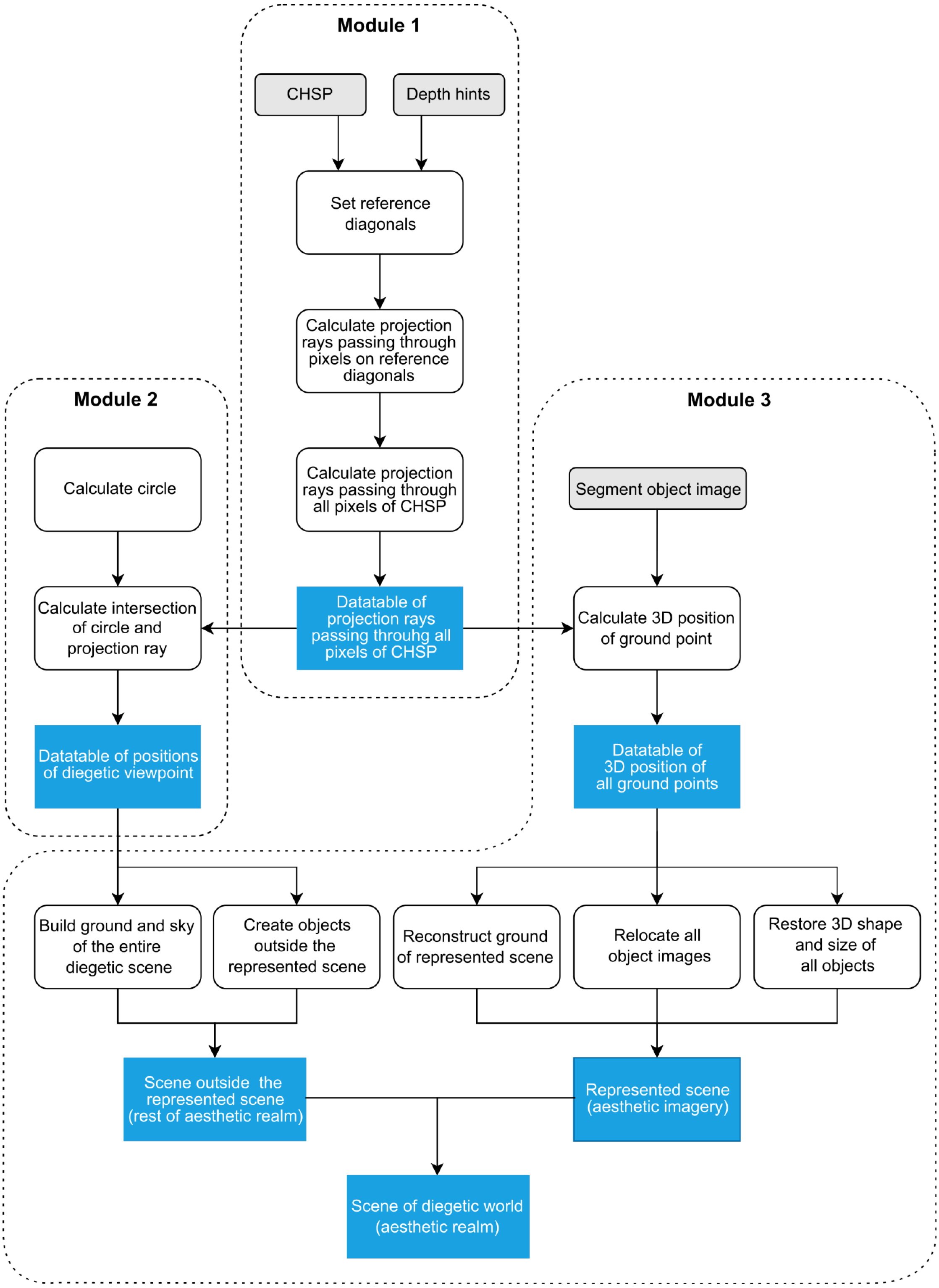

Figure 7.1.1: Overview of the algorithm design for simulating the diegetic experience

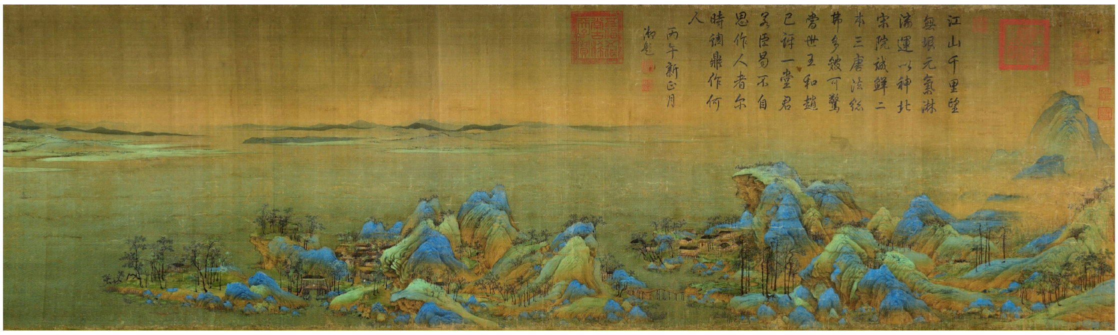

Figure 7.1.2: A Thousand Miles of Rivers and Mountains. Wang The Palace Museum, Beijing

of [Figure 5.2.2].The output is the data-table of the positions and directions of the diegetic viewpoints corresponding to all the points of the CHSP.

Module 3 is to construct the diegetic scene. According to the inference in 2.3, the diegetic scene contains two parts: the represented scene in the unrolled frame and the scene outside the represented scene that is not depicted in the CHSP. The construction of the diegetic scene is divided into two tasks: reconstructing the represented scene corresponding to the image in the unrolled frame and constructing the scene outside the diegetic scene. In the former task, the geometric correlation in [Figure 6.2.2] is applied to calculate the 3D positions of the points in the represented scene corresponding to all the points of the CHSP.And then the scene represented in the unrolled frame is reconstructed based on these data. In the latter task, a method of "random generation" is applied to create an uncertain scene based on the data from **Module 1** and **Module 2**. The output of this module is a 3D model of the represented scene and a generation method of the scene outside the represented scene.

In the development of this algorithm, the painting of *A Thousand Miles of Rivers and Mountains* 千里江山图by Wang Ximeng 王希孟(1096-1119) was selected as the research object. The painting is made of silk and measures 51.5 cm in length and 1191.5 cm in width.[2] [Figure 7.1.2] It has a typical third staged composition of open fusion, in which scenes are either connected by a long bridge or flowing water, making themselves relatively independent and interconnected.

7.2 Module 1: Computation of projection rays

The purpose of this module is to calculate the directions of the projection rays passing all the points of the CHSP. According to the formation of the cell-scene image illustrated in [Figure 4.2.7], the direction of the projection ray passing through any point in the painting is related to the diagonal passing through the same point. Theoretically, the direction of this projection ray can be computed based on the parameters of the diagonal from the point to the bottom edge of the cell-scene image, such as its inclination and the depth it represents in 3D space. These parameters can be estimated from the visual structure of the object image where this diagonal is located and the size of the object image in reality. However, the problem is that not every area in the CHSP has an object image with a well-defined visual structure. This makes it impossible to apply this method directly to the calculation of the direction of the projection rays passing through each point in the CHSP. However, according to the analysis of the visual structure of the row-scene image and the movement of the diegetic viewpoint in 6.4, the inclinations of the diagonals in

[2] The image demonstrated here is the first 1/6 part of this CHSP

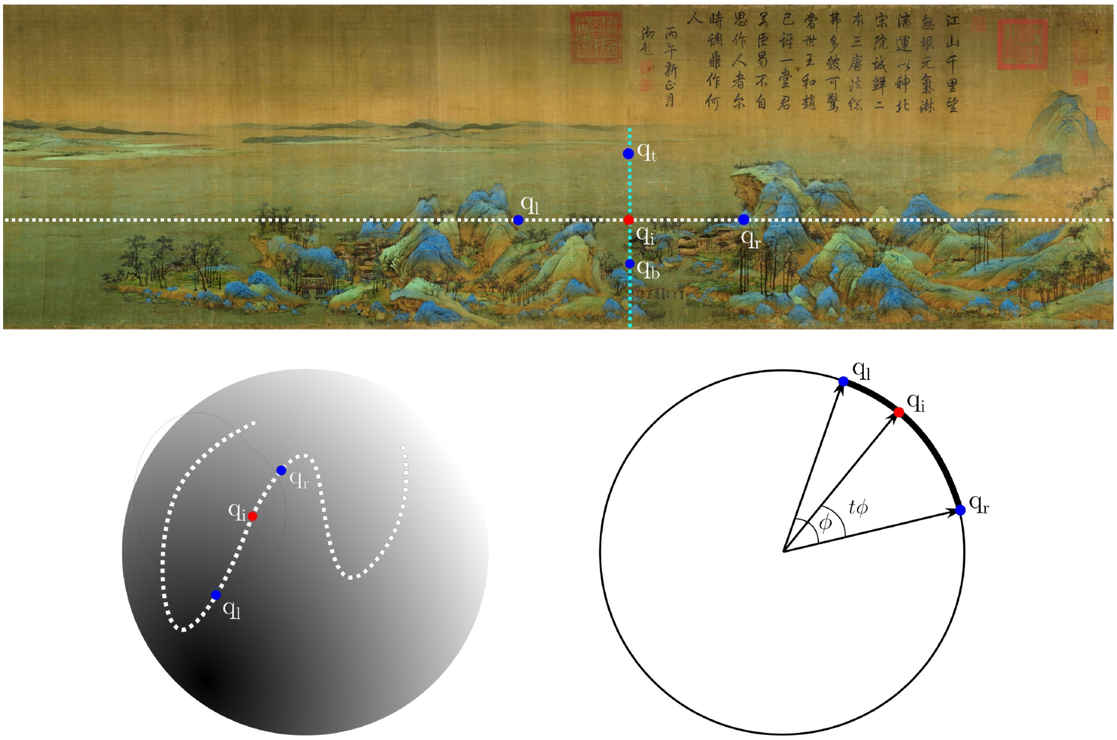

Figure 7.2.1: Calculation method for the projection ray

the third staged CHSP vary gradually in the horizontal direction. This change means that the directions of projection rays change gradually horizontally. Then, if we place the unit vectors of the projection rays passing through all points of a horizontal line in a 3D coordinate system, the connection of these vectors' ends would be a smooth curve on a sphere. [Figure 7.2.1]At this point, for any point (red node) on the horizontal line, the direction of the projection ray passing through it can be obtained by calculating the Spherical Linear Interpolation (*SLerp*) of the directions of the projection rays passing through its two left and right neighboring reference nodes (blue nodes).

Based on this hypothesis, I designed three steps to compute the directions of the projection rays passing through all points on the CHSP.

7.2.1 Set reference diagonals to provide the neighboring nodes

The method of this step is to select diagonals that can be used to calculate the direction of the projection ray, and then extend, copy and offset them to make every point of CHSP lie between two of them.

First, select reference diagonals from object images. To simplify the calculation, only the diagonals representing lines perpendicular to the canvas in 3D space are selected, such as the lateral edge of the cubic building or the connection of the mountains' shoulders, as the red diagonals show in [Figure 7.2.2].The parameters of such diagonal can be estimated according to the actual size of the objects in reality. And based on these parameters and the geometrical relationship in [Figure 4.2.7], the direction of the projection ray passing through the point on such diagonal can be computed.Therefore, the directions of the projection rays passing the points

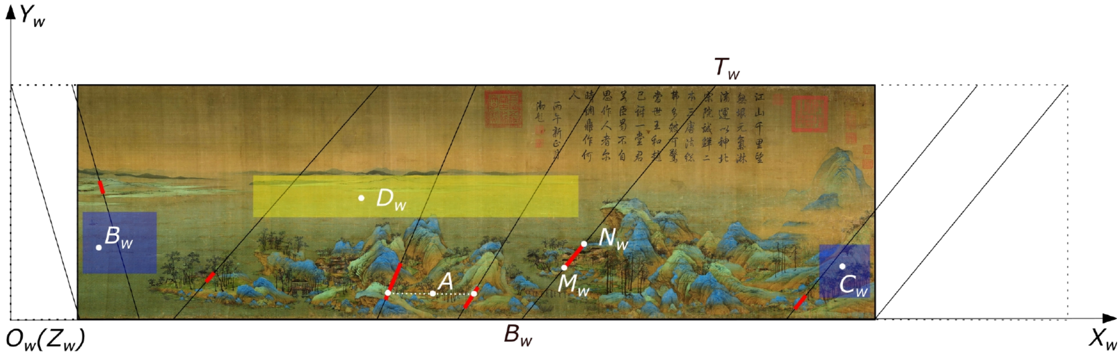

Figure 7.2.2: Draw reference diagonals

on two adjacent diagonals can be used as neighboring reference nodes for computing the direction of the projection ray passing any point between them.

Second, extend, duplicate and offset the extended diagonals. As restricted by the length and amount of diagonals selected, not all points in the CHSP can be sandwiched between two red diagonals. For example, the points in the blue and yellow areas. Therefore, two extra operations are needed to make these points lie between two depth diagonals.

a. Extend selected diagonals to the top and bottom edges of the CHSP

This operation aims to make the points in the gap section (e.g., the yellow area) lie between two diagonals.

First, I calculate the intersections of the diagonal and the top and bottom edges of the CHSP. I establish a 2D coordinate system $O_pX_pY_p$ with the origin O_p located at the lower-left corner of the CHSP. Axes X_p and Y_p respectively overlap the horizontal and vertical directions of the painting. [Figure 7.2.2]Assume that $M_p = (x_{p,m}, y_{p,m})$ and $N_p = (x_{p,n}, y_{p,n})$ are two ends of an arbitrary selected diagonal in [figure 7.2.2], then the intersection of the diagonal and the upper edge $T_p = (x_{p,t}, y_{p,t})$ and the intersection of the diagonal and the lower edge $B_p = (x_{p,b}, y_{p,b})$ can be calculated as,

$$\begin{bmatrix} x_{p,t} \\ y_{p,t} \end{bmatrix} = \begin{bmatrix} \frac{x_{p,n}h - x_{p,m}h - x_{p,n}y_{p,m} + x_{p,m}y_{p,n}}{y_{p,n} - y_{p,m}} \\ h \end{bmatrix} \tag{7.1}$$

$$\begin{bmatrix} x_{p,b} \\ y_{p,b} \end{bmatrix} = \begin{bmatrix} \frac{x_{p,m}y_{p,n} - x_{p,n}y_{p,m}}{y_{p,n} - y_{p,m}} \\ 0 \end{bmatrix} \tag{7.2}$$

Second, I make a connection between these intersections. As analyzed in 4.2.2, the ground image with unclear visual structure, such as a water image, will be perceived to have the same structural feature as the object image placed on it with a clear visual structure. Then, if only a part of the column-scene image has object images with a clear visual structure, the inclination of the visual structure of this part will be perceived as the unified inclination of the entire column-scene image. In this sense, the diagonals passing through all the points on the extended diagonal should have the same inclination. And they should overlap with the extended diagonal.

b. Duplicate the outermost extended diagonals and offset them to the left and right sides

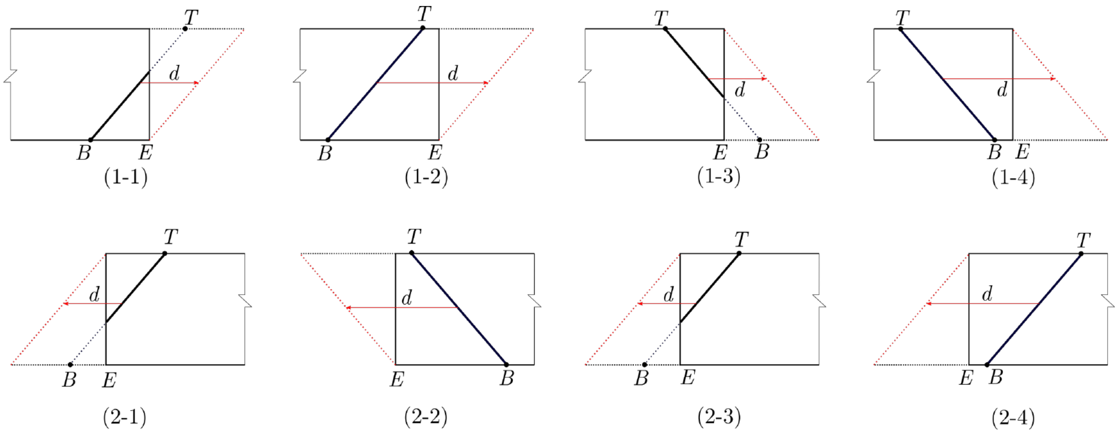

Figure 7.2.3: Cases of offsetting the outermost extended diagonals

This measure aims to make the points in the triangle areas (yellow areas) lie between two diagonals.[3]

The key to implementing this method is to figure out the offset direction and distance to offset the duplicated extended diagonal. The figures in [Figure 7.2.3] show 8 possible cases of offsetting extended diagonals. Assume that $T = (x_t, h)$ and $B = (x_b, 0)$ are the intersections of an arbitrary outmost extended diagonal and the upper and lower edges of the CHSP, and $E = (x_e, 0)$ is the horizontal coordinate of the nearest vertical edge. Then the offset direction and distance d of this diagonal can be decided by judging the following conditions.

- **Case 1-1 or 1-2**. That is, when $x_t > x_b$, the extended diagonal should be offset to the left with distance d, $d = x_t - x_b$.
- **Case 1-3 or 1-4**. That is, when $x_t < x_b$, the extended diagonal should be offset to the left with distance d, $d = x_b - x_e$.
- **Case 2-1 or 2-2**. That is, when $x_t > x_b$, the extended diagonal should be offset to the right with distance d, $d = x_e - x_b$.
- **Case 2-3 or 2-4**. That is, when $x_t < x_b$, the extended diagonal should be offset to the right with distance d, $d = x_e - x_t$.

Through these two additional operations, each point of the CHSP lies between the two extended diagonals, so that the projection ray passing it can be computed by *SLerp*.

7.2.2 Calculate the directions of projection rays passing through the points on the extended diagonals as the neighboring nodes

The method of this step is to compute the 3D positions of the points on the extended diagonals, and then calculate the directions of the projection rays passing through them based on the geometric relationships in [Figure 4.2.7].

[3]This measure is a compromise approach when there is no inclination information available outside the two ends of the CHSP. The premise of doing so is assuming that the diagonals passing through the points in these two areas have the same inclinations as the outermost extended diagonals. Besides that, it is also reasonable to consider the CHSP as a loop, then the points in the two triangle areas would lie between the two outermost extended diagonals.

The first is to calculate the x-value of the 3D position of each point on the extended diagonal. I select the column-scene image where an extended diagonal is located and establish a coordinate system $O_cX_cY_cZ_c$. Assuming that the point $P'_c = (x_{c,p'}, 0, z_{c,p'})$ in [Figure 7.2.4-left] is the represented point of P_c in 3D space, $R_c = (x_{c,r'}, 0, 0)$ is the intersection of the the extended diagonal and axis X_c, then P_cR_c is the projection of P'_cR_c onto the CHSP. As the diagonals selected represent the lines that are perpendicular to the CHSP in 3D space, then $x_{c,p'} = O_cR_c$. Given two arbitrary points $M_c = (x_{c,m}, y_{c,m}, 0)$ and $N_c = (x_{c,n}, y_{c,n}, 0)$ on the extended diagonal, then the X-coordinate of R_c can be calculated by, [Figure 7.2.4-left]

$$x_{c,p'} = x_{c,r'} = \frac{x_{c,m}y_{c,n} - x_{c,n}y_{c,m}}{y_{c,n} - y_{c,m}}. \tag{7.3}$$

The second is to estimate the y-value of the 3D position of each point on the extended diagonal. I select two segments A_cB_c and C_cD_c that lie at different heights of the diagonal and estimate their represented lengths $A'_cB'_c$ and $C'_cD'_c$ in 3D space. The estimation method is to find an object with an estimable depth at the same height as A_cB_c or C_cD_c, and use its actual depth as a reference to estimate the represented length of A_cB_c or C_cD_c in 3D space. For example, the lateral edge of the house near EF in [Figure 7.2.4-right] is estimated to be 5 meters long in 3D space according to the common dimension of the such house in reality. A_cB_c is approximately three times as long as EF, so the depth represented by A_cB_c in 3D space is approximately 15 meters. According to our definition of the depth conversion ratio (the depth in 3D space represented by a diagonal with unit length) in 6.1.2, the corresponding depth conversion ratios r_1 of A_cB_c and r_2 of C_cD_c would be $r1 = d_1/A_cB_c$, $r_2 = d_2/C_cD_c$. From the analysis of the features of the column-scene image in the third stage in 6.3.3, we have known that the depth conversion ratios in the cell-scene images that compose this column-scene image increase evenly and continuously along the vertical direction. Then the change ratio k of the depth conversion ratio in this column-scene image will be, $k = (r_2 - r_1)/B_cC_c$. Given an arbitrary point $P_c = (x_{c,p}, y_{c,p}, 0)$ on the extended diagonal, then the depth conversion ratio r_p of the cell-scene image where P_c is located will be, $r_cp = r_1 + (P_cR_c - A_cR_c) \times k$. The depth in the 3D space represented by P_cR_c is, $d_p = r_p \times P_cR_c$. As P'_cR_c is vertical to the X-axis, the $z_{c,p'} = P'_cR_c = d_p$.

The third is to calculate the direction of the projection ray passing through each point on the extension line based on its coordinates in the painting and the 3D space. Since the coordinates of the $P'_c = (O_cR_c, 0, d_p)$ in 3D space have been obtained, the projection ray passing through P_c is $\vec{v}_p = P_c - P'_c$. Then, the unit direction vector $\vec{u_p}$ of this projection ray is $\vec{u}_p = (P_c - P'_c)/|P_c - P'_c|$.

By repeating this method, the directions of the projection rays passing through all points on the extended diagonals can be obtained.

7.2.3 Calculate the direction of the projection rays passing through all points on the CHSP

The method of this step is to determine the two neighboring points of each point in the CHSP, and then to use the directions of the projection rays passing through the neighboring points as the reference nodes of the *SLerp* to calculate the direction of the projection ray passing through each point.

The first is to get the neighboring points for each point of the CHSP. This step needs to be done in the coordinate system $O_wX_wY_wZ_w$. I draw a line l parallel to axis X_w through an arbitrary point $P_w = (x_{w,p}, y_{w,p}, 0)$ on the CHSP. The intersections L_w and R_w of line l and two neighboring extended diagonals of P_w are its two neighboring points. [Figure 7.2.5]As the directions of the projection rays passing through all the points on the extended diagonals

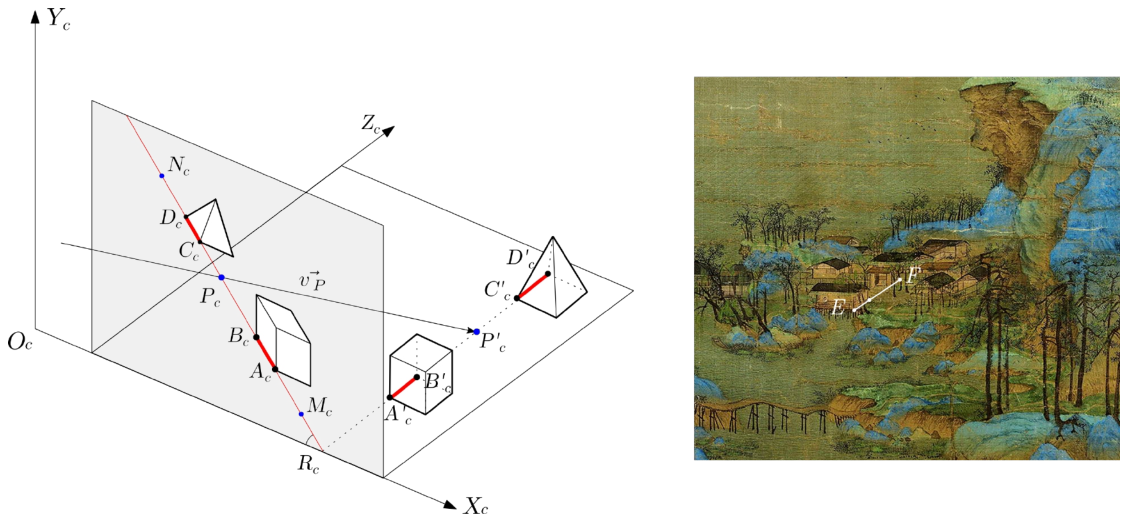

Figure 7.2.4: Select and estimate diagonals at different heights

have been calculated in Step 2, the directions of the projection rays passing through these two neighboring points can then be retrieved from the result of **Step 2**.

The second is to calculate the direction of the projection rays passing through each point in the CHSP by *SLerp*. According to the hypothesis at the beginning of this section, the directions of the projection rays passing through all points on a horizontal line change gradually. Assume that $\vec{u}_{w,l}$ and $\vec{u}_{w,r}$ are respective the direction vectors of the projection rays passing through L_w and R_w. Then, the direction vector $\vec{u}_{w,p}$ of the projection ray passing through P_w can be obtained by calculating the *SLerp* of $\vec{u}_{w,l}$ and $\vec{u}_{w,r}$.

$$\vec{u}_{w,p} = Slerp(\vec{u}_{w,l}, \vec{u}_{w,r}; t) = \frac{sin((1-t)\theta)}{sin\theta}\vec{u}_{w,l} + \frac{sin(t\theta)}{sin\theta}\vec{u}_{w,r}, \tag{7.4}$$

where t is the ratio of two distances from P_w to its two reference nodes L_w and R_w, and $0 \leqslant t \leqslant 1$. θ is the angle between $\vec{u}_{w,l}$ and $\vec{u}_{w,r}$, $\theta = cos^{-}1(\vec{u}_{w,l} \cdot \vec{u}_{w,r})$.

It should be noted that this method is the preferred method to calculate the direction of the projection ray passing any point in the CHSP. However, when $\vec{u}_{w,l}$ and $\vec{u}_{w,r}$ have the same direction or are very close to each other, $sin\theta$ will be computed as 0. In this case, this formula is invalid. To avoid the situation, $sin\theta$ should be calculated before computing *SLerp*. If it equals 0, the formula of *SLerp* should be replaced by the formula of Normalized Linear Interpolation (*NLerp*) to compute the direction of the projection ray passing this point.

$$\vec{u}_{w,p} = NLerp(\vec{u}_{w,l}, \vec{u}_{w,r}; t) = \frac{(1-t)\vec{u}_{w,l} + t\vec{u}_{w,r}}{|(1-t)\vec{u}_{w,l} + t\vec{u}_{w,r}|} \tag{7.5}$$

where t is the ratio of P_wL_w to P_wR_w, and $0 \leqslant t \leqslant 1$. At this point, the difference between the results of *SLerp* and *NLerp* is not significant.

By repeating this method above, the directions of projection rays passing through all points of the CHSP can be obtained.

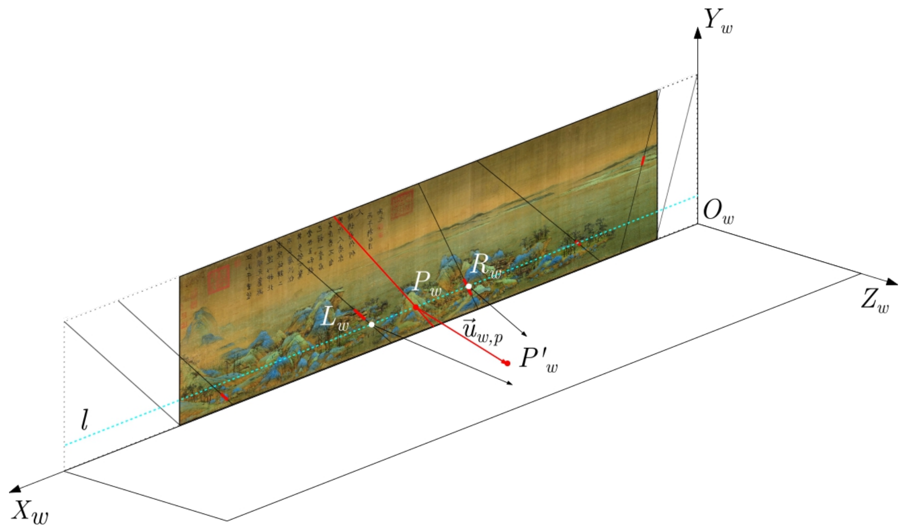

Figure 7.2.5: Calculate the direction of the projection ray passing through an arbitrary point of the CHSP

7.3 Module 2: Calculation of the position and orientation of diegetic viewpoint

According to the inference in 5.2.2, the viewer's diegetic viewpoint lies at the intersection of the projection ray and a particular circle and has the opposite direction of the projection ray. As the directions of the projection rays passing through all the points of the CHSP have already been obtained in *Module 2*, then the purpose of this module is to calculate this particular circle corresponding to each point, as well as the intersections of this circle and the projection ray.

7.3.1 Compute the circle

The method of this step is to calculate the radius and the center of the circle corresponding to each point on the CHSP, based on the geometric relations in [Figure 5.2.2].

First is the calculation of the radius r of the circle. This computation has to be done in the 3D coordinate system $O_cX_cY_cZ_c$. I make a line perpendicular to the ground through $P_c = (x_{c,p}, y_{c,p}, 0)$. [Figure 7.3.1] This line intersects the upper and lower edges of the image at $T_c = (x_{c,p}, h, 0)$ (h is the height of the CHSP) and $B_c = (x_{c,p}, 0, 0)$ respectively. According to the coincidental situation assumed in 5.2.2, the circle where the diegetic viewpoint is located should pass through T_c and B_c, and the inscribed angle α corresponds to chord T_cB_c equals to the vertical angle of the painter's FoV. We use C_c as the center of the circle which intersects with the projection ray $\overrightarrow{P'_vV_c}$ passing through P_c and denote U_c and V_c are their intersections. Next, I draw a vertical line from C_c to the chord T_cB_c intersecting with chord T_cB_c at $E_c = (x_{c,p}, 2/h, 0)$. According to the inscribed angle theorem, $\angle T_cV_cB_c = \angle T_cO_cE_c = \alpha$. Therefore, the radius r of the circle is $r = C_cT_c = T_cB_c sin\alpha/2 = hsin\alpha/2$, $C_cE_c = T_cB_c cot\alpha/2$.

The second is the calculation of the position of the circle's center C_c. Assume that, P'_c is the intersection of the ground and the projection ray passing through P_c, $\vec{n}$ is the normal vector of the ground and $I_c = (1, 0, 1)$ is a point on the plane of $O_cX_cZ_c$. Then the coordinate of intersection P'_c can be calculated by combining the equations of projection ray and ground plane

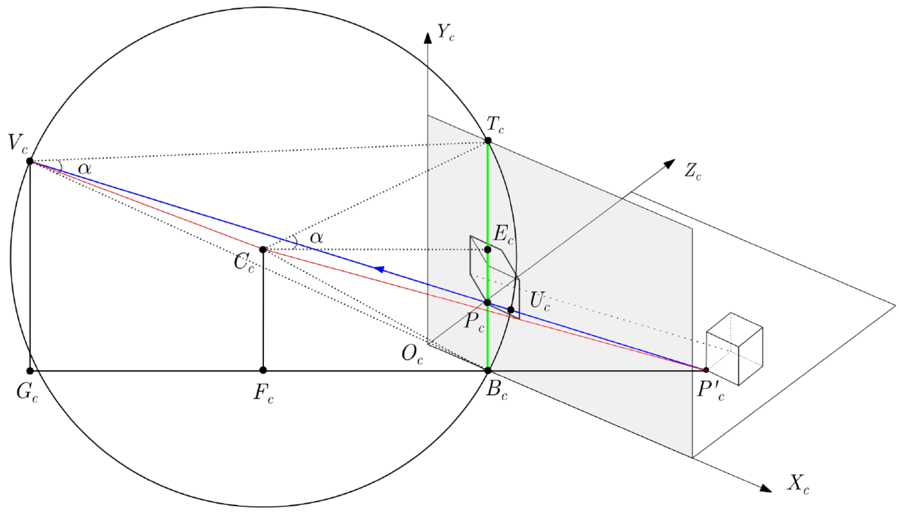

Figure 7.3.1: Calculate the position of the diegetic viewpoint

as follows,

$$\begin{cases} P'_c = P_c + d\vec{u}_{c,p} \\ (P_c - I_c) \cdot \vec{n} = 0 \end{cases} \tag{7.6}$$

Then,

$$P'_c = P_c + \frac{(I_c - P_c) \cdot \vec{n}}{\vec{u}_{c,p} \cdot \vec{n}} \vec{u}_{c,p} \tag{7.7}$$

where $\vec{u}_{c,p}$ is the unit vector of the projection ray passing through P_c that can be retrieved from the result data of Module 1. We vertically project V_c and C_c on the ground respectively at G_c and F_c. Since the chord T_cB_c is perpendicular to the ground, the circle C_c is also vertical to the ground. Then both B_c and F_c lie on the $G_cP'_c$. As C_cE_c is perpendicular to T_cB_c and both of them are on the same plane, then the $\overrightarrow{E_cC_c} = \overrightarrow{B_cF_c} = E_cC_c(\overrightarrow{P'_cB_c}/|\overrightarrow{P'_cB_c}|)$. Then the coordinate of the circle centre C_c can be obtained by $C_c = \overrightarrow{E_cC_c} + E_c$.

7.3.2 Compute the intersection of circle and projection ray

The computation of this step is to construct a system of equations for the intersections of the circle and the projection rays. According to the assumptions in the previous step, the position of V_c is where the diegetic viewpoint is located. [Figure 7.3.1]Assume that $\vec{m} = (a, b, c)$ is the unit normal vector of the circle. As T_cB_c is perpendicular to the ground plane, then the ground's unit normal vector $\vec{n}$ is the unit vector of $\overrightarrow{T_cB_c}$. Since $\vec{m}$ is perpendicular to T_cB_c and Q_cP_c, $\vec{m} = \vec{n} \times \vec{u}_{c,p}$. Then the circle can be expressed as the following system of equations based on the obtained coordinate of the circle center $C_c = (x_{c,c}, y_{c,c}, z_{c,c})$.

$$\begin{cases} (x - x_{c,c})^2 + (y - y_{c,c})^2 + (z - z_{c,c})^2 = r^2 \\ a(x - x_{c,c}) + b(y - y_{c,c}) + c(z - y_{c,c}) = 0 \end{cases} \tag{7.8}$$

And the projection ray can be expressed with $\vec{u}_{c,p} = (e, f, g)$ as,

$$\frac{x - x_{c,p}}{e} = \frac{y - y_{c,p}}{f} = \frac{z - z_{c,p}}{g} \tag{7.9}$$

Combing [Equation 7.8] and [Equation 7.9], we get the coordinates of V_c and U_c, of which V_c is one of the results with a negative z-value.

By repeating these two steps, the positions of the diegetic viewpoint corresponding to all points of the CHSP can be obtained.

7.4 Module 3: Construct the diegetic scene

This module aims to construct the scene of the diegetic world. According to the inference in 2.3, the diegetic scene can be represented as a sphere that expands outward infinitely with the viewer at the center. [Figure 2.3.3] Its scene contains two parts. One is the represented scene which is depicted in the unrolled frame. The other is the scene outside the represented scene, which corresponds to the rest of the diegetic world that has not been shown in the unrolled frame. Accordingly, the construction of the scene of the diegetic scene consists of two tasks. One is the reconstruction of the represented scene based on the image in the unrolled frame. The other is the construction of the scene outside the represented scene.

7.4.1 Reconstruction of the represented scene

The reconstruction of the represented scene generally follows a "reverse engineering" approach, that is, rebuilding the represented scene based on the reverse process of the CHSP formation. It is completed in three steps.

7.4.1.1 Reconstruct ground plane

The method of this step is to calculate the 3D positions of all points on the ground image and then move these points to their corresponding 3D positions.

The first is to calculate the 3D positions of all points on the ground image. I replace the CHSP in the coordinate system of $O_wX_wY_wZ_w$ with the ground image that has the same frame as the CHSP and define this modified coordinate system as $O_gX_gY_gZ_g$. [Figure 7.4.1]In this new coordinate system, the positions of the points on the ground image are the same as they were in the coordinate system of $O_wX_wY_wZ_w$. Assume that $\vec{u}_{g,p}$ is the unit vector of the projection ray passing through an arbitrary point P_g in the ground image, and P'_g is the projected point of P_g in the 3D space, the position of P'_g can be acquired by calculating the intersection of the projection ray and the ground plane through the same method described in [Equation 7.6] and [Equation 7.7]. By repeating this method, the positions of the projected points corresponding to all points of the ground image can be got.

The second is to move the points of the ground image to their corresponding 3D positions.[4] The ground plane is then reconstructed.

It should be noted that the ground plane in the represented scene corresponds to the displayed part of the CHSP in the unrolled frame. Therefore, for reconstructing the corresponding ground plane of the represented scene, only the points of the ground image that appear in the unrolled frame need to be moved.

[4]The moving method will be explained in Chapter 8.

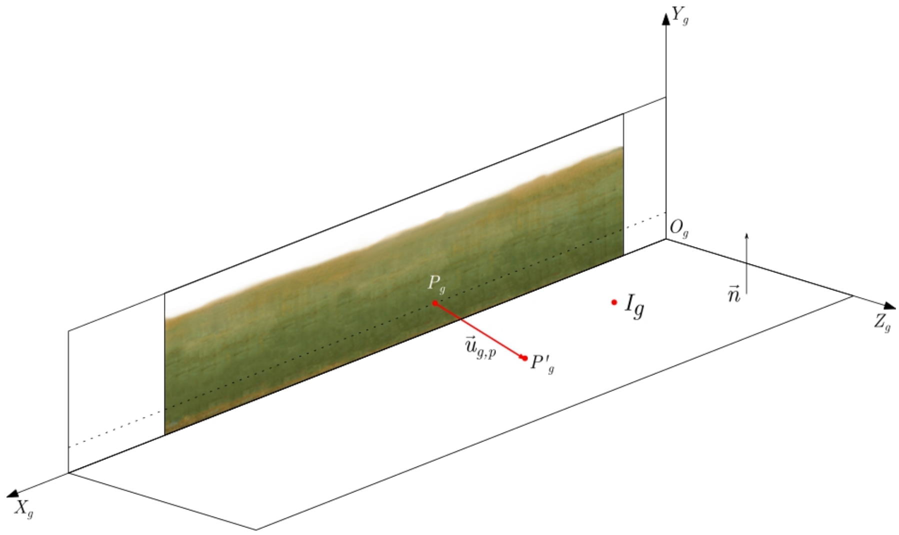

Figure 7.4.1: Reconstruction of the ground plane of the represented scene

7.4.1.2 Relocate object images

According to the analysis of the spatial structure of the cell-scene in 6.2.1, objects are placed directly on the ground plane. That means each object has a point of contact with the ground. And this ground-contact point is the position of the object in 3D space. Therefore, the method of repositioning object images is to calculate the 3D positions of the object's ground-contact point, and then place the segmented object images in their corresponding 3D positions.

The first is to calculate the 3D position of the object's ground-contact point. As the ground-contact point is both on the object and on the ground, then if we can get the corresponding 2D position of the ground-contact point on the ground image, then we can query its 3D position from the results of **Step 1**based on this 2D position.[5] Therefore, the key to calculating the 3D position of the object's ground-contact point is to accurately obtain its corresponding 2D position on the ground image.

Normally, an object has multiple ground-contact points. Here I select the foremost ground-contact point, or the vertical projection of the object's foremost point on the ground if the foremost part of the object does not directly contact the ground, as the computing target. [Figure 7.4.2] To accurately get the projection of this ground-contact point, I stroke the visual structure of the object image, as exemplified in [Figure 7.4.2-left]. Then, the projection of the ground-contact point can be easily get according to the structural vertex, as F_w (red dot) shows.

After specifying the ground-contact point, I return the 3D coordinate system $O_wX_wY_wZ_w$, from which I obtain the corresponding 2D position of this ground-contact point on the CHSP. Since I replaced the CHSP with a ground image of the same size when constructing the coordinate system of $O_gX_gY_gZ_g$, the coordinates of the points on the ground image are the same as they are in the coordinate system of $O_wX_wY_wZ_w$. Accordingly, the coordinates of this 2D position obtained above on the CHSP are equal to its coordinates in $O_gX_gY_gZ_g$. At last, I retrieve the 3D coordinates corresponding to this 2D position from the result of **Step 1**. This 3D coordinate is the 3D position of the needed ground contact point. By repeating the method, the 3D positions of the ground-contact points of all object images are obtained.

[5]The 3D positions of all points of the ground image have been obtained in **Step 1**.

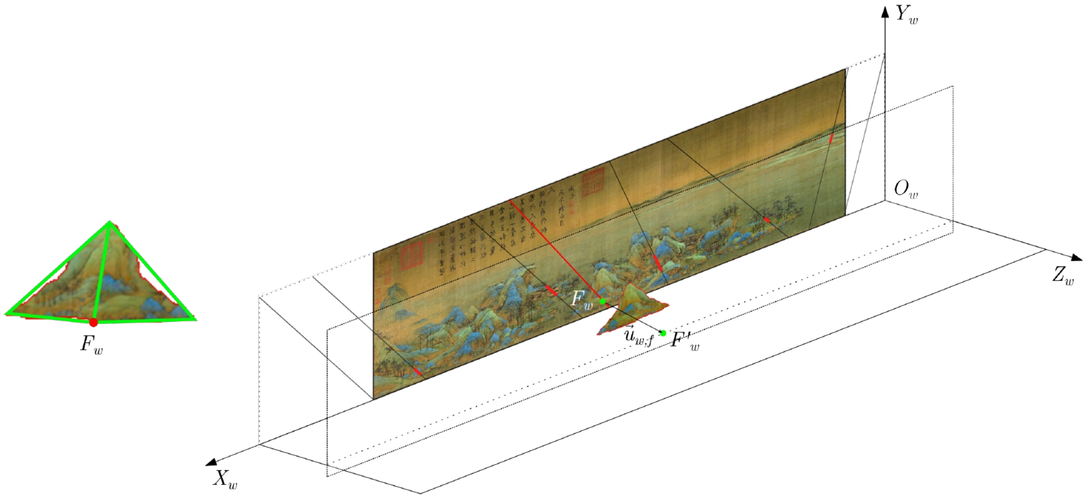

Figure 7.4.2: Relocate object image

The second is to segment object images from the CHSP and then relocate them to their ground contact point in 3D space. I separate all object images from the CHSP, complement their occluded parts, and place them in the same frame as the CHSP. Then, I move the object image so that the projection of the ground contact point on the object image coincides with the ground contact point in the 3D space. In this way, the repositioning of object images is completed.

7.4.1.3 Restore the 3D shape of object

According to the formation, cell-scene image analyzed in 4.2.2, the 3D shape of the object was "compressed" into 2D in the first stage by following a process that is the same as the oblique projection. In this process, the vertices were moved at different distances along the same projection direction. The restoration process of the object's 3D shape is the reversal process of this "compression". That is to move the points of the object image to their respective distances along the projection direction. Since the projection direction has been obtained (the same direction as the projection ray passing through the ground contact point), the key to restoring the 3D shape of the object is to calculate the distance of each vertex moved along the projection direction. I design four steps to complete these tasks.

The first is to calculate the difference between the maximum and minimum distances at which the object's vertices moved. I replace the ground image in the coordinate system of $O_gX_gY_gZ_g$ with an object image attached to a full-size frame and set this modified coordinate system as $X_oY_oZ_oO_o$. Assuming that F_o and R_o are two points on the object image, which respectively represent the nearest ground contact F'_o and the farthest ground contact R'_o, $\vec{u}_{o,f}$ is the direction of the projection ray passing through F_o. By querying the results of **Module 1** and **Module 2**, we obtain the value of $\vec{u}_{o,f}$ and the coordinates of F'_o and R'_o. Then, the difference between the distances that N_p and F_o were moved along $\vec{u}_{o,f}$ in the projection transformation

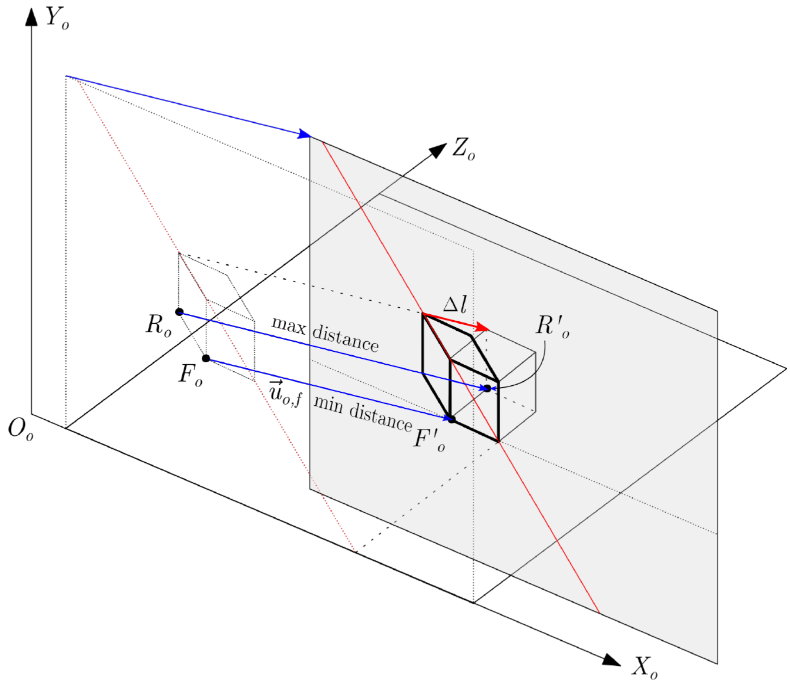

Figure 7.4.3: Calculate the maximum difference in moving distance

is $\Delta l = F_oF'_o - R_oR'_o$. This value is the distance difference we need. [Figure 7.4.3]

The second is to distinguish the vertices' relative depths to the nearest ground contact point F_o. To do this, the height-map of each object image is needed. I make it manually by setting the points in different depth regions with different grayscales.[6] Firstly, I frame regions on it in the sequence from far to near and record the points in the framed region each time, as the mountain image shows in [Figure 7.4.4].Secondly, I set the grayscale of the point with the least number of frames to 0 and the point with the most number of frames to 255. For an arbitrary point of the image, I calculate its grayscale by the times it was framed. Assuming that m is the total number of times the figure was framed and n is the time this point was framed, then its grayscale g should be $g = 255 - 256n/m$. In this way, I compute all points' grayscales and then make the height-map of an object.

The third is to calculate the relative moving distance of each object's vertex. According to the above calculations, Δl is the difference between the distances that the object's nearest ground contact point F_o and the farthest ground contact point R_o moved. As their corresponding points on the height-map were set to 0 and 255, Δl corresponds to 256 greyscales. Then, the relative moving distance that one grayscale represents is $\Delta s = \Delta l/256$. Given an arbitrary vertex P_o in the object, the relative distance it was moved is $\Delta l_p = \Delta s \times g$, where g is the difference in greyscale between F_o and R_o on the height-map.

Fourth it is to restore the object's 3D shape. As the projection rays passing through all points of an object image have the same direction with projection ray passing through F_o, then the

[6] Depth differentiation is an important research area in computer vision research. For photographs, depth differentiation can generally be performed with the help of a uniform pattern, such as the structure of the graph, the change of color, etc. However, the depth representation in Chinese paintings is complex, and even the same painting does not use a unified way to represent depth. Therefore, it is difficult to find an absolute rule. For this reason, I use the current manual approach.

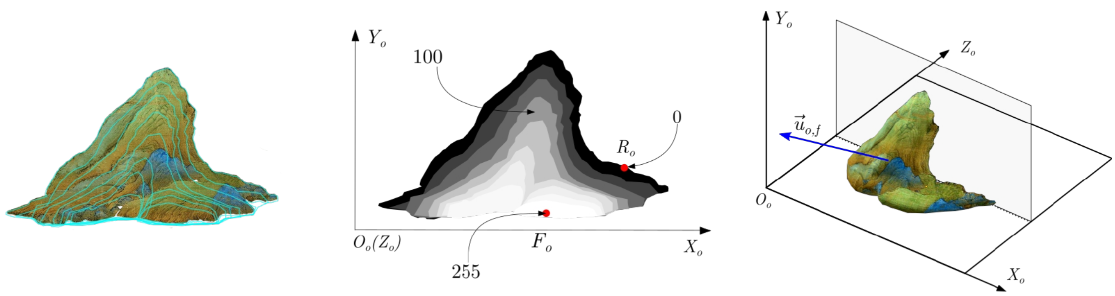

Figure 7.4.4: Calculate the maximum difference in moving distance

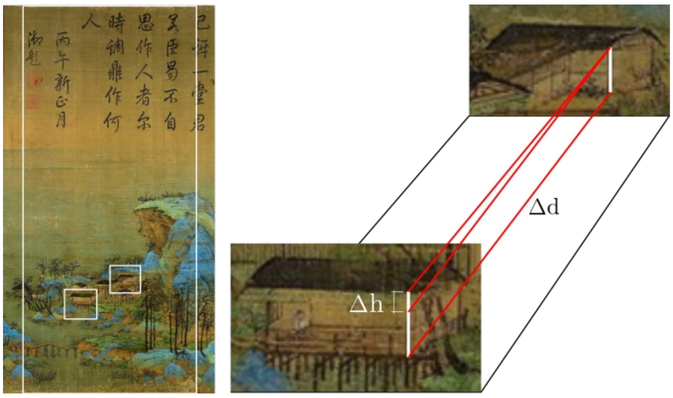

Figure 7.4.5: Calculate diminished scale

vector $\vec{v}_{o,p}$ that the arbitrary vertex P_o needs to be moved is $\vec{v}_{o,p} = -\vec{u}_{o,p}\Delta l_p$. By this method, I calculate the vectors that all vertices need to be moved and then moved these vertices along their vectors to restore the preliminary 3D shape of the object.

However, this 3D shape is not the final shape of the object. Since the image of the object has been moved to its 3D position of the ground-contact point in Step 2, the object would appear smaller than it does on the CHSP if it is seen from the diegetic viewpoint. Therefore, the restored 3D objects need to be scaled up.

Here I follow the assumption in 5.2.2 that abstracting the viewer's eyes as a camera with a fixed FoV.Then the size of the object in the vision is inversely proportional to its distance from the diegetic viewpoint. I select two object images located at different heights in CHSP, such as the two houses in [Figure 7.4.5]. Since these two houses are of the same type, they can be treated to have the same dimensions. Assume that Δh is their height difference in CHSP and Δd is their relative distance in 3D space. Then, the ratio k of the object size in vision to its relative distance to the projection plane is $k = \Delta h / \Delta d$. The object's size in the virtual environment should be scaled up to $S = D * k$, where D is the distance from the object to the vertical projection of its corresponding diegetic viewpoint. Last, I used the object's ground-contact points as an anchor to scale the primary 3D object to its proper size.

By repeating the above steps, the 3D shapes of all the objects appearing in the unrolled frame can be restored and then the reconstruction of the represented scene can be finally completed.

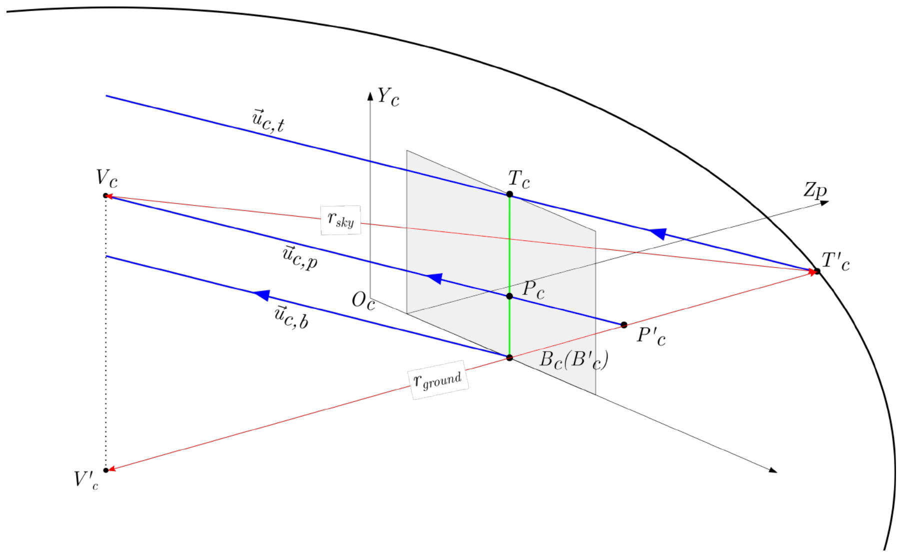

Figure 7.4.6: Centers and radius of the ground plane and sky

7.4.2 Construction of the scene outside the represented scene

The scene outside the represented scene is the rest of the scene after the spherical diegetic scene minus the represented scene corresponding to the image in the unrolled frame. It contains the complete sky and the rest part of the round ground after it minus the reconstructed ground, as well as objects. The construction of these things is completed in two steps.

7.4.2.1 Construct the ground and the sky

The method of this step is to compute the centers and radius of the round ground and spherical sky and build them based on their equations.

In a sphere-shaped diegetic scene, the ground is a roundness centered on the diegetic viewpoint's vertical projection. This round ground extends from this center to the infinite distance until it meets the sky. Therefore, the distance from this center to the junction line of the ground and sky is the radius of the round ground.

The second is to compute the radius of the ground and the sky. I return the coordinate system $O_cX_cY_cZ_c$ with a single column-scene image extracted from the CHSP. [Figure 8.4.6] Assume that V_c' is the vertical projection of the diegetic viewpoint V_c on the ground, T_c is the intersection of the vertical line (the green line) passing through P_c and the top edge of the ground image, then T_c's corresponding projected point T_c' in the 3D space is located on the edge of the ground plane. The distance from V_c' to T_c' is the radius of the round ground, where the 3D position of T_c' can be obtained by querying the results of Step 1 in 7.4.1. I use V_c' as the center and $V_c'T'c$ as the radius to build up a complete ground of the diegetic world. Then, the rest of the ground plane, excluding the reconstructed ground in 7.4.1, is the scene's ground outside the represented scene.As the ground intersects the sky at its farthest edge, the farthest endpoint of the ground is also on the sky. Then, V_cT_c' is the radius of the sphere. [Figure 7.4.6]By creating a sphere with V_c as its center and V_cT_c' as its radius, the sky can be constructed.[7]

[7]These details to create them will be talked about in the next chapter.

7.4.2.2 Create objects outside the represented scene

According to the analysis in 2.3, the scene outside the represented scene is a pure conception of the viewer. Therefore, all subjective factors of the viewer, such as his understanding of the painting, previous experiences in the landscape, and cultural literacy, would influence his conceptions of the content and layout of this scene. The objects and their layouts then may vary from person to person or even from time to time, which creates uncertainty in the scene outside the represented scene. And this uncertainty means that the scene outside the represented scene cannot be fixed into one setting or another.

As an interpreter of the CHSP, the virtual exhibition system needs to explain this uncertainty to the VR user. This means that a scene of uncertainty needs to be built. Here, I use "random generation" to realize this purpose. Specifically, it is to generate a random scene outside the represented scene for each position of the diegetic viewpoint.[8] Thus, even at the position of the same diegetic viewpoint, different VR users or the same VR user at different times will see the different scenes outside the represented scene.

Considering that the objects in this scene expanded based on the represented scene, the objects within and outside the represented scene have the same style. Therefore, it is more practical to create them by replicating the objects that had been restored and randomly arranging them on the ground outside the represented scene.[9] So that all objects can remain the same. In this way, the objects in the scene outside the represented scene can be created.

Summary

This chapter elaborates on the algorithm for computing the direction and position of the diegetic viewpoint and constructing the diegetic scene. The algorithm has three modules.

Module 1: Calculation of the projection rays of the entire image of the CHSP

This module is implemented in three steps.

Step 1: Set reference diagonals to provide the neighboring nodes

The method is to select diagonals that can be used to calculate the direction of the projection ray, and then extend, copy and offset them to make every point of CHSP lie between two of them.

Step 2: Calculate the direction of the projection rays passing through the points on the extended diagonals

The method is to compute the 3D positions of the points on the extended diagonals, and then calculate the directions of the projection rays passing through them, based on the geometric relationships in [Figure 4.2.2].

Step 3: calculate the direction of the projection rays passing through all points on the CHSP

The method is to determine the two neighboring points of each point in the CHSP, and then to use the directions of the projection rays passing through the neighboring points as the reference nodes of the *SLerp* to calculate the direction of the projection ray passing through each point.

[8] The implementation of the random generation will be talked about in the next chapter.

[9] It is also available to create these objects from scratch but with the same textures as the represented objects.

Module 2: Calculation of diegetic viewpoint's position

This module is realized in two steps.

Step 1: Compute the circle

The method is to calculate the radius and the center of the circle corresponding to each point on the CHSP, based on the geometric relations in [Figure 5.2.2].

Step 2: Compute the intersection of circle and projection ray

The method is to construct a system of equations for the intersections of the circle and the projection rays.

Module 3: Construction of the scene of the diegetic world

The construction consists of two parts: the reconstruction of the represented scene and the construction of the scene outside the represented scene.

(1). Reconstruction of represented scene

Step 1: Rebuild the ground

First, calculate the intersection of the projection ray passing through each point and the ground by establishing the equations of the two. Second, reconstruct the ground by moving the points on the ground to their corresponding positions in 3D space.

Step 2: Relocate the object images

First, query the 3D position of the foremost ground contact point of each object image from the result of Step 1. Second, move each segmented object image onto its ground-contact point.

Step 3: Recover the object's 3D shape and scale

First, calculate the max distance that the object is compressed. Second, generate a height map of the object image. Third, calculate the relative moving distance of each vertex. Fourth, move the points on the object image along their respective corresponding vectors to recover the 3D shape of this object. Fifth, calculate the size reduction ratios of the objects and recover their scales.

(2). Construction of rest scene of diegetic world

Step 1: construct the ground and sky

First, obtain the diegetic viewpoint's vertical projection as the center of the ground. Second, draw a vertical line on the CHSP passing through the point and get the intersections of this line and the upper and lower edges of the ground image. Third, query the 3D position of the upper intersection. The distance between the round center and this 3D position is the radius of the round ground. The distance between the diegetic viewpoint and this 3D position is the radius of the sphere sky. Third, use the ground center and ground radius to create a round as the complete ground of the diegetic world. Fourth, use the diegetic viewpoint as the center and the sky radius to create a sphere as the sky of the diegetic world.

Step 2 is to construct the rest scene of the diegetic world

The method is to randomly select and replicate the restored objects and randomly arrange them on the ground outside the represented scene.

Part IV

Synchronization

Chapter 8

Program implementation of virtual exhibition system

This chapter will elaborate on the program implementation for synchronously simulating the viewing experience and the diegetic experience of "touring in the CHSP" on the VR platform. First, the design of the synchronous simulation program will be briefly introduced. Second, this chapter will illustrate the computations for three necessary data: the directions of the projection rays (diegetic sightlines) passing through all points on the CHSP, the orientations and positions of the diegetic viewpoint corresponding to all points on the CHSP, the 3D positions of the points corresponding to all points on the CHSP. Third, the reconstruction of the viewing scene and the constructions of the diegetic scene will be illustrated. Last, this chapter will explain the development of the VR synchronous simulation program.

8.1 Program design of the virtual exhibition system

According to the analysis in 3.4, the diegetic experience of "touring into CHSP" has two features.First, the touring experience in the diegetic world synchronously corresponds to the viewing experience in the real world, which is embodied in two aspects. On the one hand, the diegetic scene is imagined based on the image in the unrolled frame. On the other hand, the viewer's diegetic viewpoint corresponds to his focus point, thus ensuring that the diegetic scene he sees is consistent with the orientation and position of the image presented in the unrolled frame. Second, the diegetic movement is a dynamic and viewer-directed action. The viewer needs to scroll the CHSP and move the focus point while watching it. These manipulations respectively make the image in the unfolded frame dynamically change and the position of the focus point change. Since diegetic experience corresponds to the viewing experience, the diegetic scene would change and the viewer's diegetic viewpoint would move synchronously. From this point, the viewer's experience in the diegetic scene is a dynamic movement controlled by himself. These two features are fundamental to the aesthetic experience of the CHSP, which needs to be interpreted in the exhibition system.

For this purpose, I designed a nested virtual space, integrating a viewing scene into the diegetic scene. [Figure 8.1.1]This viewing scene represents the ancient viewing place. It contains two necessary components: a viewing platform on which the VR user can stand and freely move; a real-sized virtual CHSP[1] which allows being manipulated like as a real one, like scrolling,

[1] To reduce calculations, we select 1/6 of the length of this CHSP as the test object.

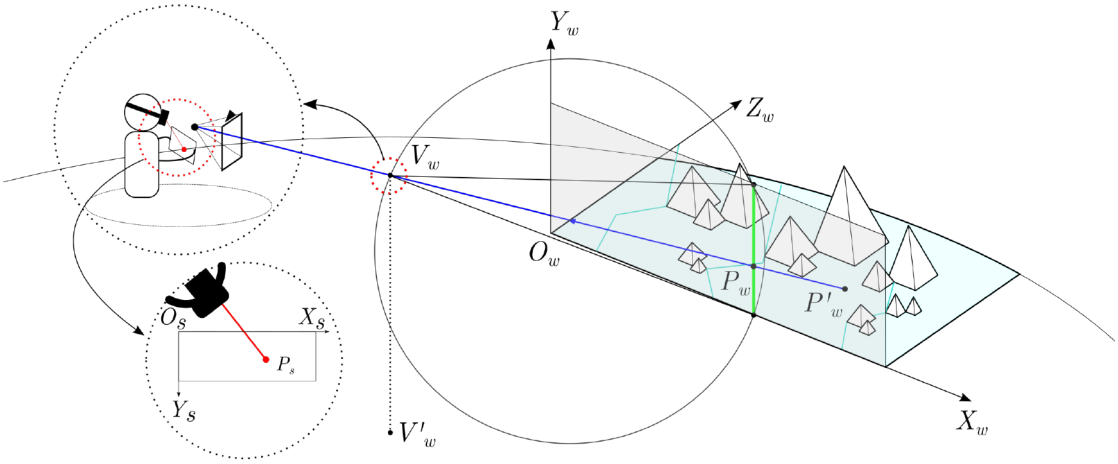

Figure 8.1.1: Design of the virtual exhibition system

stretching, etc. The diegetic scene represents the aesthetic realm. It is a spherical space have two parts: the represented scene, which corresponds to the image in the unrolled frame, the scene outside the represented scene, which is created randomly.

The viewer's experiences in the two scenes are correlated through a synchronization program. This program will monitor the position of the VR user's focus point on the virtual CHSP and then adjust the position and orientation of the entire viewing making the front of the viewing platform always face the represented scene.[2] Meanwhile, it will also monitor the displayed image in the unrolled frame when the VR user scrolls the virtual CHSP. Then, it will adjust the represented scene according to the current image and update the scene outside the represented scene according to the position of the current diegetic viewpoint in real-time.

For this synchronization program, an ideal operation would be to calculate the direction and position of the diegetic viewpoint and construct the diegetic scene in real-time based on the position of the VR user's focus point on the virtual CHSP and the image in the unrolled frame. However, in practice, two factors make this ideal operation difficult to achieve. First, data computation and scene reconstruction cannot be fully automated. Some processes, such as picking reference diagonals, separating object images, and selecting ground contact points, require VR makers' intervention. This intervention is not possible during the VR user's manipulation. Second, the current VR devices[3] cannot run the entire computational and construction process smoothly. According to the algorithm developed in Chapter 7, all the data needed to simulate the diegetic experience are related to the pixels on the CHSP. This means that high resolution will affect the operating speed of the program.[4] That ideal way is not feasible under current conditions.

For this reason, I designed a compromising approach to achieve real-time correspondence

[2]Here I chose to adjust the viewing platform rather than the VR user's viewpoint. The reason is that the VR user should be free to move on the platform. By this method, the represented scene always appears in front of the VR user in his first glimpse, which echoes the correspondence between the represented scene and the image in the unrolled frame.

[3]The equipment used for the virtual exhibition system is Oculus Quest. It is a head-mounted VR device with six degrees of freedom. Its two handles can simulate the interaction with CHSP in reality. The hardware includes Qualcomm Snapdragon 835 CPU, 4G DDR memory, and Adreno 540 GPU.

[4]The highest resolution of the target CHSP is 39,974 * 1,600 pixels. For this test, I selected a section of the target CHSP and adjusted its resolution to 800*234.

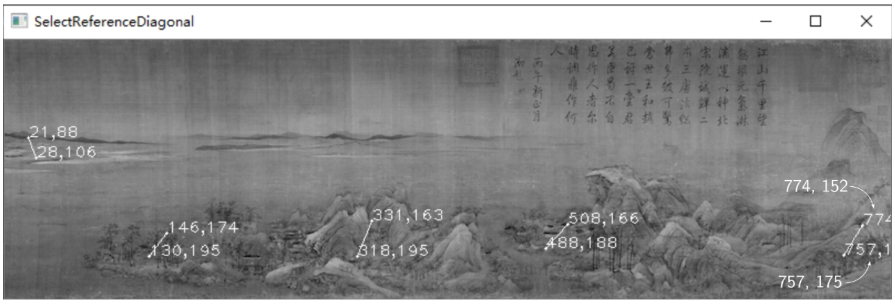

Figure 8.2.1: Select reference diagonals

between the two experiences. That is to pre-calculate the necessary data and pre-build the represented scene corresponding to the fully expanded CHSP, and then, use a synchronization program to retrieve the required data and control the displayed part of the represented scene. This idea is implemented by the following three steps.

Step 1: calculate fundamental data

This step aims to compute three sets of data. the directions of the projection rays passing through all pixels on the CHSP, the positions of the projected points in 3D space corresponding to all pixels on the CHSP, and the positions of the viewer's viewpoint in the scene of the diegetic world corresponding to all pixels on the CHSP. These data are computed according to the computing methods developed in 7.2, 7.3, and 7.4. All results are output as data-tables.

Step 2: reconstruct represented scene

This step aims to reconstruct the represented scene of the whole CHSP. This reconstruction is based on the reconstruction method in 7.4.1. The result of the construction is output as a 3D model.

Step 3: develop synchronizing program

The synchronous program consists of three parts: Module for synchronizing the movement of the VR user's focus point on the virtual CHSP and the movement of the viewing platform, Module for synchronizing the image in the unfolded frame and the displayed part of the reconstructed represented scene, Module for randomly generating the scene outside the reconstructed scene. This program is presented as a VR application.

8.2 Data preparation

8.2.1 Computation of the directions of the projection rays passing through all pixels of the CHSP

This calculation is performed according to the method for calculating the *SLerp* corresponding to each projection ray passing through each point developed in 7.2. This method is implemented in three steps.

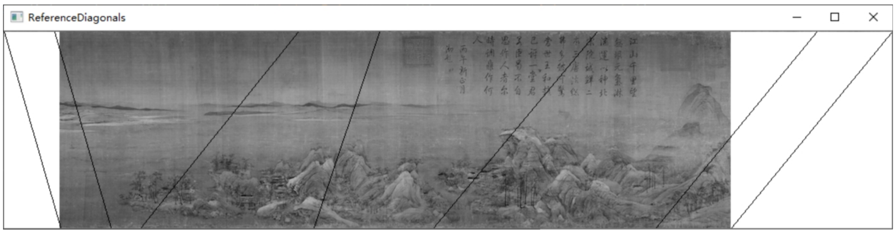

Figure 8.2.2: Duplicate and offset the outermost extended diagonal

8.2.1.1 Set reference diagonals to make each point of the CHSP lie between two of them

According to the method designed in 7.2.1, this step includes three operations.

First, select diagonals from object images. Here I used the method of mouse click (`cv.setMouseCallback()`) to get the two endpoints of the diagonal. Then, I used the method of drawing line(`cv.line(img, (ix, iy), (x, y), (255, 0, 0), thickness=1)`) to connect these two endpoints. [Figure 9.2.1]

Second, extend the selected diagonals to the upper and lower edges of the CHSP and get the coordinates of all points on the line. To complete this operation, I firstly calculated the intersections of the diagonal and the upper and lower horizontal edges according to the [Equations 7.1 and 7.2]. Secondly, I adopted *Bresenham's line algorithm*[5] to draw a line between these intersections and get the coordinates of all the pixels on this line.

Third, duplicate the outermost extended diagonals and offset them. In this operation, I firstly determine the direction in which the outermost extended diagonal needs to be moved and calculate the distance to be offset, according to the cases shown in [Figure 7.2.3].secondly, I duplicated and offset the two outmost extended diagonals by copying their coordinates and modifying the pixels' x-coordinates. Assume x_0 is the x-coordinate of an arbitrary pixel on the left extended diagonal, and d_l is the distance it needs to be offset. Then, the x-coordinate x_1 of this pixel on the left duplicated extended diagonal should be modified as $x_l = x_0 - d_l$. If x_l is the x-coordinate of an arbitrary pixel on the right extended diagonal, then the x-coordinate x_1 of this pixel on the left duplicated extended diagonal should be modified as should be $x_r = x_1 + d_r$.

The image in [Figure 8.2.2] shows the final result of these three operations based on the selected diagonals in [Figure 8.2.1].

8.2.1.2 Calculate the directions of the projection rays passing through the points on the reference diagonals

According to the method in 7.2.2, this step needs to be done by three operations.

First, select segments of different heights from the extended diagonals. Here I still use the method of mouse click to select these diagonal segments. The image in [Figure 8.2.3-right] shows the selection results.

Second, estimate the length of these selected diagonal segments and enter input these lengths. According to the estimation method in 7.2.2, the actual depths should be estimated based on the object image, like a house or man figure, near the selected segments. But the unit of length we use in reality is different from the unit of length calculated by the computer. Therefore, before input, we first need to convert the estimated length in meters to a length in pixels. The method finds an object image near a diagonal segment and compares the size of that object image with

[5] https://www.cs.helsinki.fi/group/goa/mallinnus/lines/bresenh.html

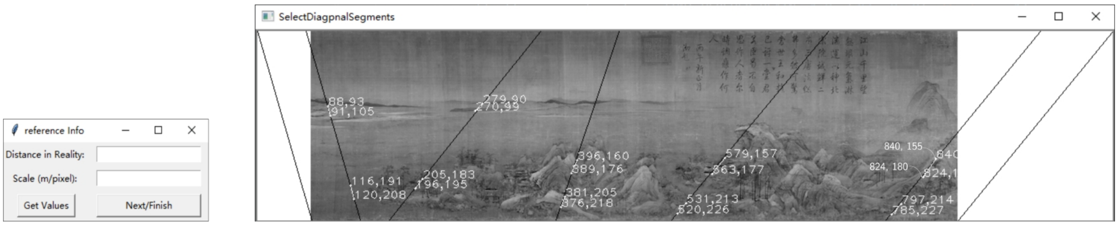

Figure 8.2.3: Interface and Selection of diagonal segments at different heights

Line	x1	y1	x2	y2	Scale
Segement 1	88	93	91	105	35
Segement 2	116	191	120	208	5
Segement 3	279	90	270	99	30
Segement 4	205	183	196	195	4
Segement 5	381	205	376	218	1.7
Segement 6	396	160	389	176	1.8
Segement 7	579	157	563	177	2
Segement 8	531	213	520	226	1.5
Segement 9	840	155	824	180	4
Segement 10	797	214	785	227	4.5

Table 8.1: Input depth and scale

its size in reality. For example, the actual depth of the house in [Figure 8.2.3] is 5 meters, and it occupies 5 pixels on the image. Then in this part of the image, one-pixel length represents 1m in reality. The unit convention scale for this diagonal is 1. By this method, I estimated the converting scale for each diagonal segment. Secondly, I used *Tkinter*[6] to make a panel for inputting the estimated depths and scales (`tkinter.Entry(tkinter.Tk(className='Reference Info'), textvariable=tkinter.StringVar())`). [Figure 8.2.3-left]The input depths and scales of selected diagonal segments are shown in [Table 8.1].

The third is to calculate the directions of the projection rays passing through all the pixels on the selected diagonals. I imported the data of each diagonal segment of [Table 8.1] into the formulas listed in 7.2.2 and compute the unit direction vector of the projection ray passing each point on the extended diagonal. The result is saved as *Datatable-diagonal* with the following data structure: $[x_w, y_w, i, j, k]$, where (x_w, y_w) is the coordinate of each pixel in the coordinate system $O_wX_wY_w$. (i, j, k) is the unit direction vector of the projection ray.

8.2.1.3 Calculate the directions of the projection rays passing through all points on the CHSP

According to the method designed in 7.2.3, this step includes two operations.

The first is to find the two ends of the interval where each point of the CHSP is located. I firstly selected the points (data row) that have the same y-value as $P_w = (x_{w,p}, y_{w,p}, z_{w,p})$ from the *Datatable-diagonal*. These points are the ones in the same row as P. Secondly, I used these points as endpoints to divide the entire row into intervals (`Interval(start, end)`). Last, I

[6]https://docs.python.org/fr/3/library/tkinter.html

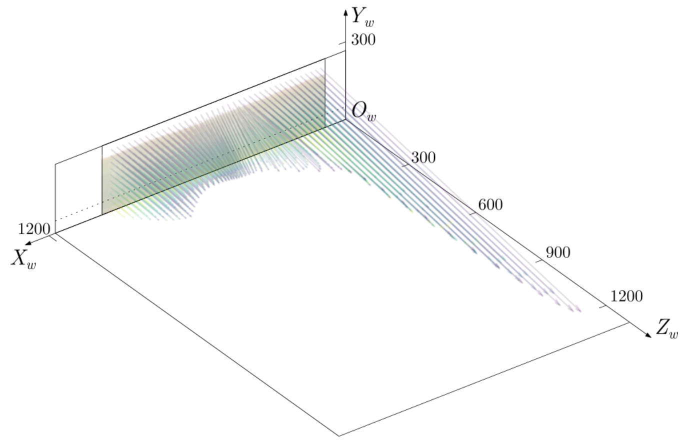

Figure 8.2.4: Projection rays passing through all pixels

decide which interval the P_w lies in, according to $x_{w,p}$, and obtain the two endpoints of this interval.

The second is to calculate the direction of the projection ray passing through each point of the CHSP. As analyzed in 7.2.3, when the interval length tends to be 0, the *SLerp* formula is invalid and the computation has to switch to *NLerp*. Therefore, it is necessary to conduct a conditional judgment ahead of the computation. So, I queried the unit vectors of two endpoints of the interval from *Datatable-diagonal*. Then, I calculated the angle θ between the projection rays passing through the two intervals' endpoints and then decide which method to use. If θ is not equal to 0, we use the function coded according to [Equation 7.4]. If θ is equal to 0, we use the function coded according [Equation 7.5]. The return results of these two functions are the directions of the projection rays of the arbitrary P_w.

The third is to iterate all the pixels on the CHSP and repeat the above calculations. I stored all computed results in the *Datatable-ray*, and set its structure as the same as *Datatable-diagonal's*. The diagram in [Figure 8.2.4] illustrates the directions of the projection rays passing through all pixels according to the data in [Table 8.1].

8.2.2 Computation of the positions of the diegetic viewpoint corresponding to all pixels of the CHSP

This calculation was done by following the method for calculating the intersection of each projection ray and the special circle corresponding to each point of CHSP developed in 7.3. As developed, this method needs to be implemented in two steps.

The first is to build the equation of the circle corresponding to each point of the CHSP. I imported the projection ray from *Datatable-ray* into the geometrical equations in 7.3.1 row by row to compute the circle's center and radius corresponding to each point of the CHSP.

The second is to calculate each intersection of the projection ray and the circle. I iterated the projection ray from *Datatable-ray* and the circle's center and radius and imported each of them into the equation system of projection ray and circle in 7.3.2. The result with a negative z-value is the position of the diegetic viewpoint.

I saved all the results in a *Datatable-viewpoint* and set its structure as follows: $[x_w, y_w, x_v, y_v, z_v]$. Of which, (x_w, y_w) is the coordinate of each pixel in the coordinate system $O_wX_wY_w$, and (x_v, y_v, z_v) is the coordinate of the diegetic viewpoint. The scatter in [Figure 8.2.5] illustrates

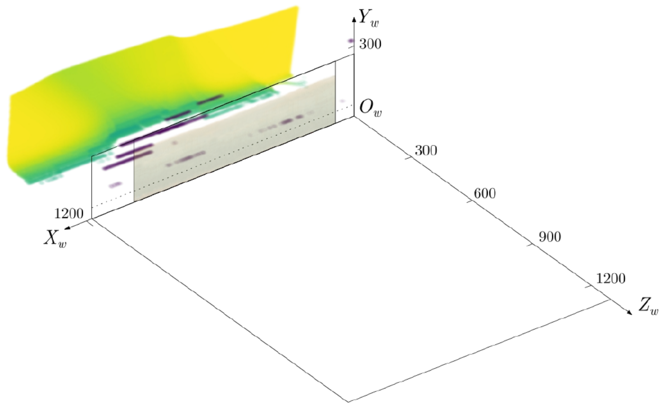

Figure 8.2.5: Positions of the diegetic viewpoints corresponds to all pixels

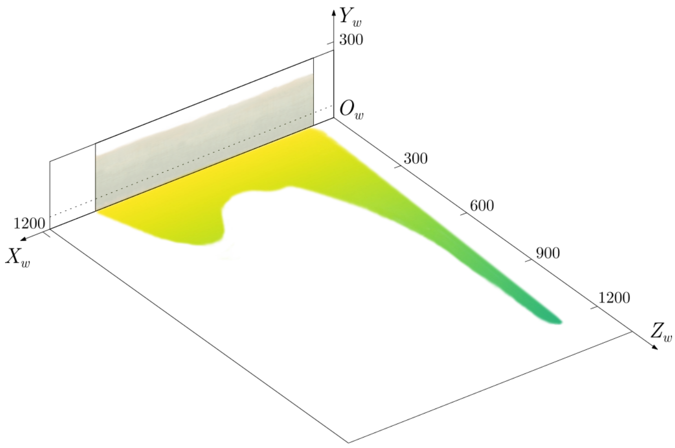

Figure 8.2.6: Positions of the projected points correspond to all pixels of the ground image

the results calculated according to the data in *Datatable-ray*.

8.2.3 Computation of the 3D positions of all the points in the ground image

This computation was accomplished by following the method for calculating the intersection of each projection ray and ground developed in 7.4.1. I iterated *Datatable-ray* row by row and input each row data into the geometrical equations in 7.4.1. The results were stored in a *Datatable-ground*. Its structure was set as follows: $[x_w, y_w, x_g, y_g, z_g]$. Of which, (x_w, y_w) is the coordinate of each pixel in the coordinate system $O_wX_wY_w$, and (x_g, y_g, z_g) is the coordinate of the projected point in 3D space. The scatter in [Figure 8.2.6] illustrates the results of *Datatable-ground* computed according to the *Datatable-ray*.

8.3 Model preparation

According to the program design in 8.1.1, two scenes need to be built in advance: the viewing scene and the represented scene corresponding to the whole CHSP.

8.3.1 Construction of the viewing scene

The viewing scene is used to simulate the viewing experience of the viewer in the real world. It should include at least two necessary components: a viewing platform and a virtual CHSP.

8.3.1.1 Build the viewing platform

In ancient China, the places to watch CHSP were diverse. It could be a private study room or a pavilion for gathering. So there are many references for building such a viewing place. However, considering that this viewing scene is placed in the diegetic scene, it is recommended to keep its style the same as the diegetic scene. Therefore, rebuilding a building or pavilion in CHSP may be a solution. The rebuild method is the general 3D modeling method. This building process won't be covered again here.

8.3.1.2 Create a virtual CHSP

The virtual CHSP is designed to simulate the viewer's manipulation of the CHSP. Therefore, this virtual CHSP not only requires the same shape as the actual CHSP but also needs to be able to simulate the responses after various manipulations, such as scrolling, stretching, etc. Accordingly, this task is completed in two steps.

(1). Create a virtual CHSP

I firstly imported a scanned full-size CHSP into Unity and created two cylinders as two rollers[7]. Secondly, I created a *ScrollMask* as an image mask to control the displayed part of the CHSP. At two ends of this *ScrollMask*, I attached the two rollers.[Figure 8.3.1-left] I set that only the part that is covered by *ScrollMask* can be displayed. So a transparent material was given to this *ScrollMask*. And a Stencil Structures was added to the shaders of *ScrollMask* and the image. The ref parameters of this shader's Stencils[8] were set as follows,

```
ScrollMask: Stencil{
            Ref 1
            Comp Greater
            Pass Replace}
Image: Stencil{
            Ref 2
            Comp Equal}
```

(2). Create scripts to simulate the scrolling manner

According to the manipulations in reality, the scrolling manner is double hands operation with one hand holding a roller and the other hand scrolling the other roller.

To simulate this manipulation, I first added the preset script `OVR Grabbable` from *Oculus Integration Package* to both rollers so that the VR user can grab them by pressing the `Axis1D.PrimaryHandTrigger` of either handle. [Figure 8.3.1-left]

[7] According to the analysis in 3.1. One end of the CHSP is a cylinder, the other is a slim bar. But for this test, I simplified them into two cylinders.

[8] https://docs.unity3d.com/2019.3/Documentation/Manual/SL-Stencil.html

Figure 8.3.1: Design of virtual CHSP

Second, I created two empty objects, *LeftEmpty* and *LightEmpty*, at the center of two rollers and set the hierarchy of objects in the *ScrollRig* was set as [Figure 8.3.1-right].

Thirdly, I coded a script to simulate the scrolling manner. In this script, I set two cases.

Case 1: when the `Axis1D.PrimaryHandTrigger` of either handle is pressed alone (`leftRoller.grabbedBy == leftGrabber && rightRoller.grabbedBy != rightGrabber`), the script will use `transform.SetParent` to modify the hierarchy of the objects in the *ScrollRig* as [Figure 8.3.2-middle], so that the complete *ScrollRig* will be moved along the held roller.

Case 2: when one handle's `Axis1D.PrimaryHandTrigger` is being pressed and the other handle's `Axis1D.PrimaryIndexTrigger` is being pressed (`OVRInput.Get(OVRInput.Button.SecondaryIndexTrigger) == true`), this script will use the same method to adjust the hierarchy of the objects in the *ScrollRig* as [Figure 8.3.2-right]. At this moment, the VR user can pull one end of the *ScrollMask* with the roller at this side moving along.

Fourthly, I coded a script to edit the length of the displayed part of the CHSP in **Case 2** dynamically. I set the *ScrollMask* to keep its length equal to the distance between the two handles when the distance between the two handles is between 0 and the length of the image. But when adjusting the length of *ScrollMask*, I did not directly set its length to the distance between the two handles, but used the distance between the two handles as the x-value in the scaling operator `transdorm.localScale` to scale the LeftEmpty or the RightEmpty in the roller being held. That is to code as [9]

```
float currentDistance = Vector3.Distance(currentLeftRollerPos,
    currentRightRollerPos);
leftEmpty.transform.localScale = new Vector3(currentDistance * 4, 1, 1);
```

In this way, when moving one handle to stretch the *ScrollMask* from one side, the other side of the *ScrollMask* being held will remain fixed. Besides, it is also necessary to make the roller at the scrolling synchronously follow the movement of the stretched edge of the *ScrollMask*. So I kept the position and orientation of this roller to be consistent with the moving edge by,

```
Bounds scrollMaskBoundsLocal = scrollMask.transform.GetComponent<MeshFilter>().
    mesh.bounds;
Vector3 rightBoundsLocal = new Vector3(scrollMaskBoundsLocal.max.x, 0,
    scrollMaskBoundsLocal.extents.y);
rightRoller.transform.position = scrollMask.transform.TransformPoint(
    rightBoundsLocal);
rightRoller.transform.rotation = scrollMask.transform.rotation;
```

[9]These codes are particularly designed for stretching to the right. Stretching to the left will be applied in the same method.

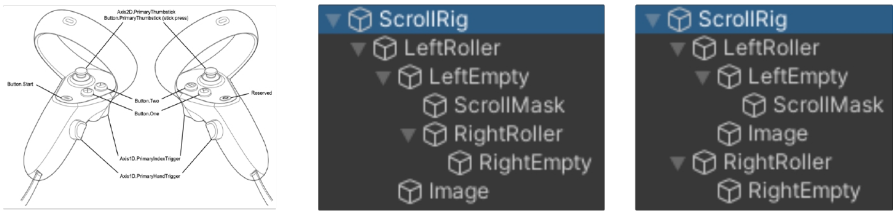

Figure 8.3.2: Setting of the objects' hierarchies in Case 1 and Case 2.

8.3.2 Reconstruction of the represented scene of the whole CHSP

In current image-based 3D reconstruction projects, the results are usually presented in two ways: point cloud models and sculptural meshes. Considering that the first way will occupy a large amount of memory during the operation and may cause the program to lag, I choose the second way as the presentation of the reconstruction results. The represented scene was built separately.

8.3.2.1 Reconstruct the ground of the represented scene

This reconstruction is conducted by three steps developed in **Step 1** of 7.4.1.

(1). Segment ground image

Here, I used the graphic editor to segment the ground image from the CHSP and complement its occluded part and place it in a frame that has the same size as the CHSP[10].

(2). Create mesh of ground image

Firstly, I imported the segmented ground image and built a planar mesh in the blender environment with the same dimensions as the ground image. I assigned the ground image as the texture to the material of this mesh and change its shader node as shown in [Figure 8.3.3]. In this way, I rendered the white space of the image transparent.

Secondly, I used the operator of `bpy.ops.mesh.subdivide(number_cuts)` in the Blender-bpy package to divide this mesh into small faces and get the coordinates of all the vertex of this divided mesh. The number of times the mesh is divided (the value of `number_cuts`) relates to the number of vertexes that can be manipulated when reshaping this mesh.[11] Theoretically, the number of vertices after division should be the same as the number of pixels of the ground image. Then we can reconstruct the 3D ground plane by moving each vertex of the ground image to its corresponding 3D position. However, too much vertex can cause Blender to crash or hang[12]. Therefore, it is necessary to balance the number of divisions according to the fineness of the generated model and the processing power of the current equipment we use. Considering all factors, I set this value to 30. The image shown in [Figure 8.3.4-left] is the effect after division.

(3). Restore the shape of the ground

First, I imported the *Datatable-ground* obtained in 8.2.3 and queried the 3D positions of the points that match the coordinates of the vertices on the mesh.In practice, I queried *Datatable-*

[10] The technique of image segmentation is an important research area in development. The existing segmentation techniques are currently automated, but they do not yet meet the requirements for fine detail. Therefore, we still use the manual segmentation method.

[11] According to the blender document, after n times subdivision the number m of the vertex will be $m = ((2^n + 2)^2)/4$. See details at https://archive.blender.org/wiki/index.php/Doc:2.6/Manual/Modifiers/Generate/Subsurf/

[12] Performance Considerations: https://docs.blender.org/manual/en/2.79/modeling/modifiers/generate/subsurf.html

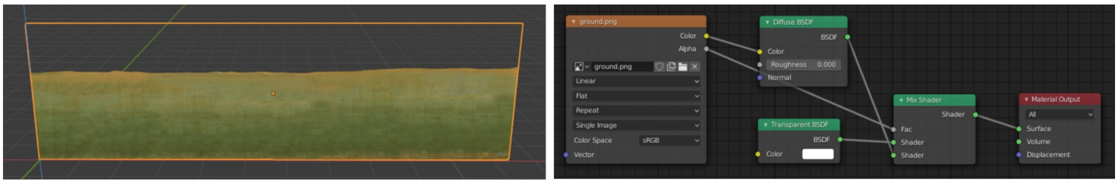

Figure 8.3.3: Planar mesh and shader nodes of the ground plane.

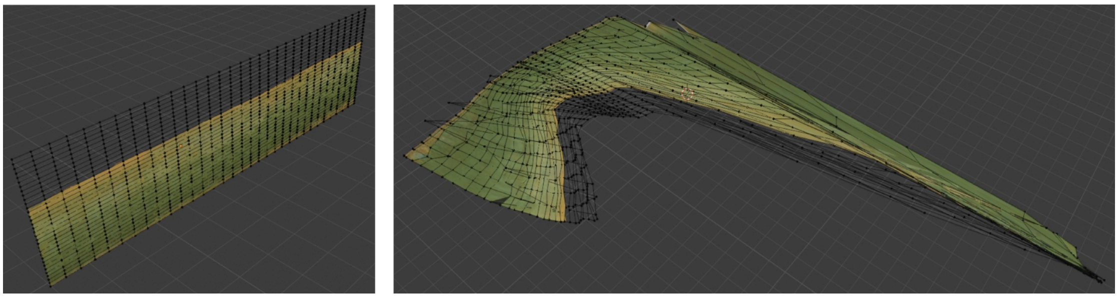

Figure 8.3.4: Rebuild the ground

ground for the row that can simultaneously satisfy $x_w = x_v$, $y_w = y_v$, where (x_v, y_v) is the local coordinate of an arbitrary vertex of the mesh. If the above equations can be satisfied, then the coordinates (x_g, y_g, z_g) in the same row are the position of the projected point corresponding to the arbitrary vertex. Theoretically, the vertices of the mesh and the rows of *Datatable-ground* are one-to-one matching if we make the number of vertices of the divided mesh the same as the number of pixels on the ground image.

Second, I used the operator of `bpy.ops.transform.translate` to move each vertex to its corresponding 3D position to complete the reconstruction of the ground plane. That is, `bpy.ops.transform.translate(value=`$(x_g, y_g, z_g) - (x_w, y_w, 0)$`)`. However, in the previous division step, I adjusted the number of divisions so that the coordinates of some vertices may not match with every row in the *Datatable-ground*. Therefore, to make these vertices also move approximately to their corresponding 3D positions, I set the `use_proportional_edit = True` of the operator and adjust the value of `proportional_size =30`. In this way, the vertices in the `proportional_size` region will be moved at proportional distances along the translated vertex. Then, those vertices that do not match any row in the *Datatable-ground* will also move to their corresponding 3D positions approximately.

In this way, I finally completed the reconstruction of the entire ground plane. [Figure 8.3.4-right]

8.3.2.2 Relocate object image

The relocation of the object image is conducted by three steps illustrated in **Step 2** of 7.4.1.2.

(1). Segment object image

Here, I also used the manual method to segment all object images from the CHSP, complement their obscured parts, and place them in a frame of the same size as the CHSP with their relative positions unmoved.

(2). Retrieve the position of the ground-contact point in 3D space

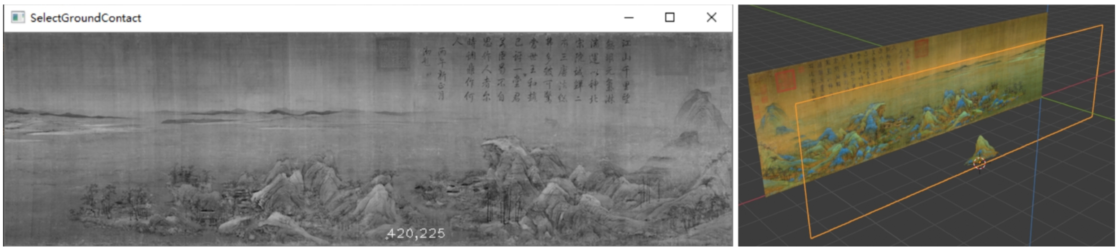

Figure 8.3.5: Relocate object images on 3D positions

First, I created an interactive panel and show the segmented object image in it. [Figure 8.3.5-left] I also used `CV.EVENT` to record the position of the click.

Second, I imported *Datatable-ground*, which records the depth data of the points on the ground image, and queried the data row that can match the coordinates of the selected point of the ground-contact point.

(3). Move the object image onto the 3D position of its ground contact point

First, I created a planar mesh in Blender's environment with the same dimension as the CHSP. Second, I imported the segmented object image, gave it to this mesh's material as a texture, and set its shader nodes as [Figure 8.3.3-right]. In this way, its blank space will be rendered transparent. Third, I applied `bpy.types.ToolSettings.transform_pivot_point = 'CURSOR'` to move the cursor to the selected projection point of the ground-contact point, and then used `bpy.ops.object.origin_set (type='ORIGIN_CURSOR', center='MEDIAN')` to move the center point of the object image to the cursor. Last, I adopted `bpy.ops.transform.translate()` to move the object image onto the 3D position of its ground-contact point. The value in this operator is the vector that the object image needs to be moved, which is `value =` $(x_g, y_g, z_g)-(x_w, y_w, 0)$. [Figure 8.3.5-right]

By repeating the **(2)** and **(3)**, I relocated all the segmented object images to their corresponding 3D positions.

8.3.2.3 Restore the 3D shape and scale of object

The restoration of the object is conducted by three steps explained in **Step 3** of 7.4.1.3. This restoration was implemented in five steps.

(1). Get the longest relative distance the object vertices move in the projection process

First, I created a new interactive interface and displayed the segmented object image. In this interface, I clicked the nearest and the further ground-contact points' corresponding 2D points on the object image with a full-sized CHSP frame and then recorded the coordinates of these two points.

Second, I imported the *Datatable-ground* and got the positions of the farthest and the nearest ground-contact points by using the query method before. And then I input these two positions into the computation method in **Step 3** of 7.4.1.3 to get the moving distances of the two points along the projection ray. This distance is the longest relative distance that the object's vertices move.

(2). Make the height map of the object image

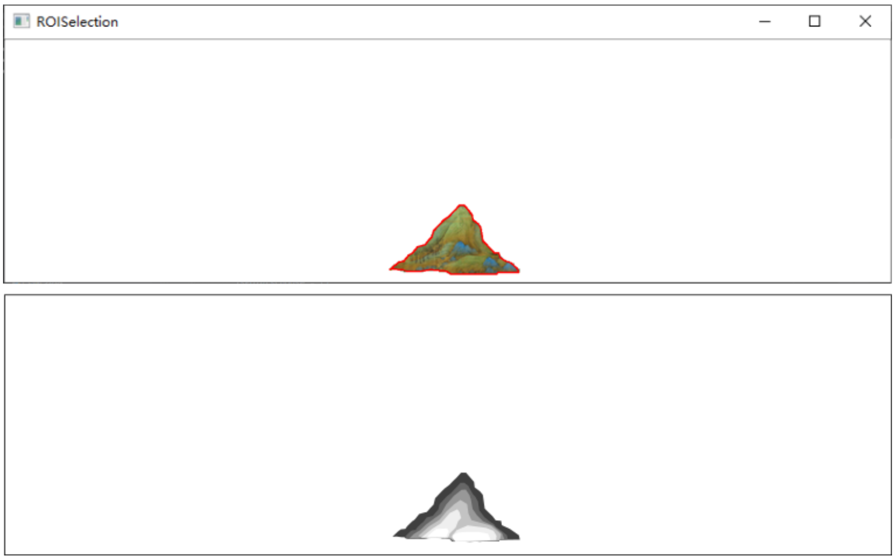

Figure 8.3.6: Generate height-map of object image

First, I converted the object image to a gray-scale image and presented it in an interactive window. Second, I framed the regions representing different depths on the object image from far to near by mouse click. [Figure 8.3.6-top]. Third, I recorded the total number of outlining times and the coordinates of the framed pixels each time. Last, I changed the gray-scale of each pixel according to the method in 7.4.1.3. The result is shown in [Figure 8.3.6-bottom].

(3). Calculate the distance that each pixel of the object image needs to move

First, I imported the height-map made in the previous step and iterate the pixels on the height-map for their the gray-scales. Second, I calculated the vector of each pixel that needs to move by inputting each row of *Datatable-grayscale* and the maximum relative moving distance of the object's vertex into the computation method in 7.4.1.3. Last, all the results of the function's return were saved into a *Datatable-vector* in the structure as $[x_w, y_w, i_m, k_m, k_m]$, where (i_m, k_m, k_m) is the pixel's moving vector.

(4). Move the vertices on the object image with their respective vectors that need to move

First, I used the operator of `bpy.ops.mesh.subdivide`[13] to divide the mesh of the object image created in **Step 2** into small faces and get the coordinates of all vertices.

Second, I queried each vertex's corresponding moving vector from the *Datatable-vector*. Similar to the situation when reconstructing the ground plane, the vertices correspond to the pixels one by one in the ideal state. Then, we can directly move the vertices by using `bpy.ops.transform.translate(value=`(i_m, k_m, k_m)`)` to complete the restoration of the object's 3D shape. However, too many vertices can cause the Blender to clash. Therefore, it is necessary to need to adjust the subdivision number. But this would make some vertices can not match any of the rows in the *Datatable-vector*. As is the case discussed in reconstructing the ground plane, we can use the proportional edit method (`propotional_size`) to move these vertices approximately. However, the object itself is relatively finer than the ground plane. Thus, it is not

[13]Here we set the `number_cuts=50`, after balancing the hardware and the needed result.

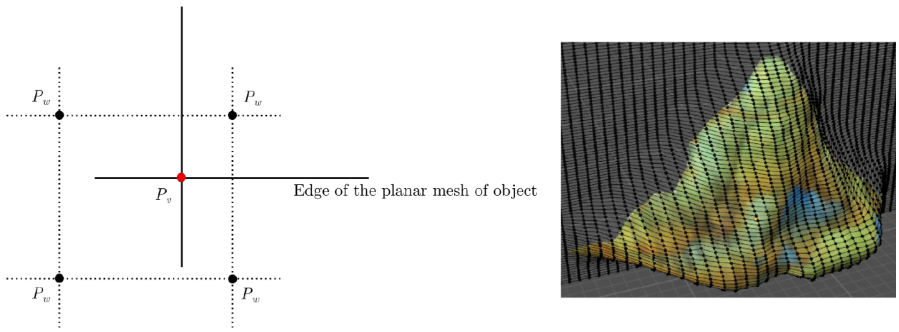

Figure 8.3.7: Restore the 3D shape of object image

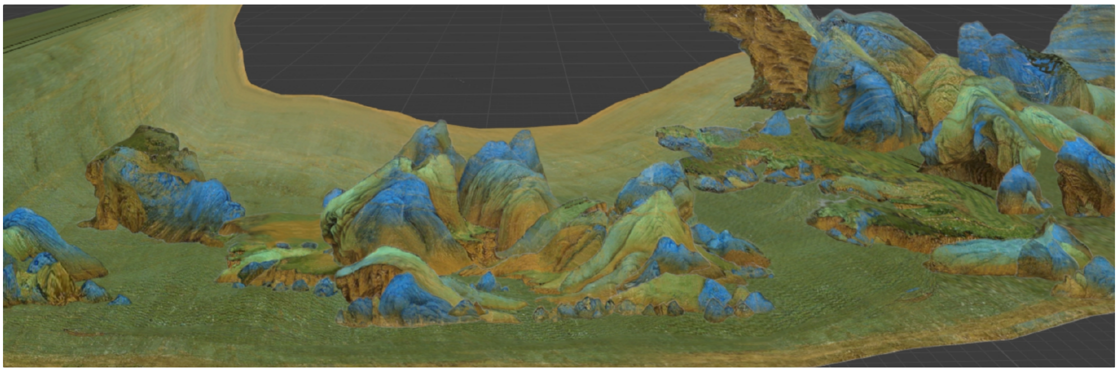

Figure 8.3.8: Reconstruction of the represented scene

suitable to be applied this method to the object image. For this reason, I adopted an approach to finding the nearest point. Assume that (x_v, y_v) is an arbitrary vertex P_v on the subdivided mesh. The specific method is to use `scipy.spatial.KDTree` to convert the pixels' coordinates to tree-shaped data, and then to find the nearest pixel from the tree by applying `scipy.spatial.KDTree.query`[14]. As shown in [Figure 8.3.7-left], assume that the red dot is the mesh's vertex and the black dots represent the pixels of the object image which all have corresponding data in the *Datatable-vector*. What needs to do is to find the nearest black point to the red point. After obtaining the nearest pixel, I queried its corresponding moving vector from the *Datatable-vector* and used this vector as the moving vector of P_v. [Figure 8.3.7-right]

Last, I moved the vertices of the mesh along their corresponding moving vectors by using `bpy.ops.transform.translate(value=`(i_m, k_m, k_m)`)` to complete the restoration of the 3D shape of the object. To make the model surface smooth, I set `use_proportional_edit = True` and `proportional_size = 3`.

By repeating the above four steps, I restored the 3D shape of all objects and finally reconstructed the represented scene corresponding to the whole CHSP. [Figure 8.3.8]

(5). Restore the scale of the object

The above data calculation and model construction were done assuming that the painting and its diegetic world have the same scale. In this situation, the viewer was simplified as a viewpoint without scale inside this miniature landscape. When the viewer enters this virtual space by VR

[14]https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.KDTree.html

device, the viewer's scale will be restored, and he will use his scale as a reference to judge the scale of his surroundings. To maintain the relative scale of the virtual scene in the viewer's vision, both the computed data and the reconstructed model need to be adjusted accordingly.

For this reason, I firstly calculated the size ratio between the CHSP and the represented scene in reality. Secondly, I selected a house in CHSP as the object. Assume that h (in meters) is its height in the painting and H is its height in reality. Then the size ratio between CHSP and the represented scene is $k = h/H$.[15] Thirdly, I adjusted the scale of the imported represented scene according to this ratio and divided each calculated value by k for their new values in the virtual environment.

8.4 Development of synchronous program

The synchronous program consists of three parts: Module for synchronizing the movement of the VR user's focus point on the virtual CHSP and the movement of the viewing platform, Module for synchronizing the image in the unfolded frame and the displayed part of the reconstructed represented scene, Module for randomly generating the scene outside the reconstructed scene.

8.4.1 Synchronization between the movement of the focus point on the virtual CHSP and the movement of the viewing platform

The method for this module is to detect the position of the VR user's focus point on the virtual CHSP, then query its corresponding diegetic viewpoint's direction and position from the computed data, and finally translate the viewing platform according to these data.

(1). Detect and get the position of the focus point in real-time

First, I added a ray-casting script to the *CameraCenterEyeAnchor*[16] in *OVRCameraRig*. This script will create a ray that emits from the center of the VR camera's frustum. I used it as the sightline of the VR user. Then, the intersection of this ray and the virtual scroll is the position of the VR user's focus point on the virtual scroll. Second, I added *RaycastHit*[17] in the script to monitor and get the position of the intersection of this ray and the virtual scroll in real-time. To ensure that the intersection of a ray and virtual scroll always stays within the unrolled frame, we set to get the coordinate of the intersection only when the ray intersects with the mask and image at the same time. [Figure 8.4.1]

(2). Query the projection ray's direction and the position of the diegetic viewpoint corresponding to the position of the focus point in real-time

I added a data-matching script to the VR program. This script select the direction vector of the corresponding projection ray from *Datatable-ray* and the position of the corresponding diegetic viewpoint from *Datatable-viewpoint* by using `datatable.Select(filter string)`[18], of which the direction of the projection ray is the direction of the diegetic viewpoint.

(3). Translate the viewing platform according to the queried direction and position in real-time

[15] For the current CHSP, I estimated the size ratio k as $k = 1/500$.

[16] CameraCenterEyeAnchor is a preset VR camera in the Oculus Integration Package. It represents the center of the VR camera's frustum.

[17] https://docs.unity3d.com/ScriptReference/Physics.Raycast.html

[18] The dynamic synchronization programs were develop in C#. And `datatable.select` is a method from .NET. https://docs.microsoft.com/en-us/dotnet/api/system.data.datatable.select?view=net-5.0

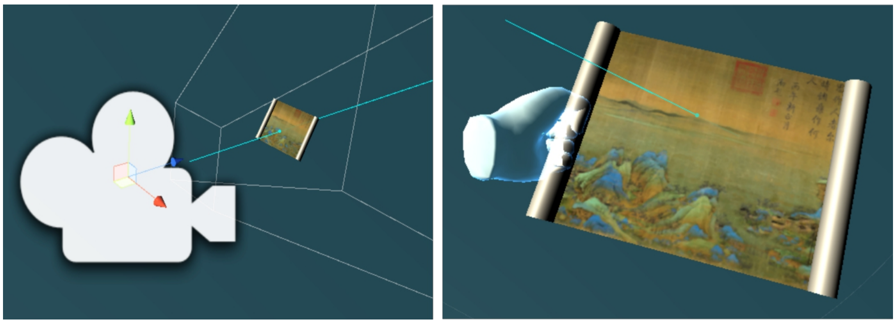

Figure 8.4.1: Get the position of focus point on the virtual CHSP

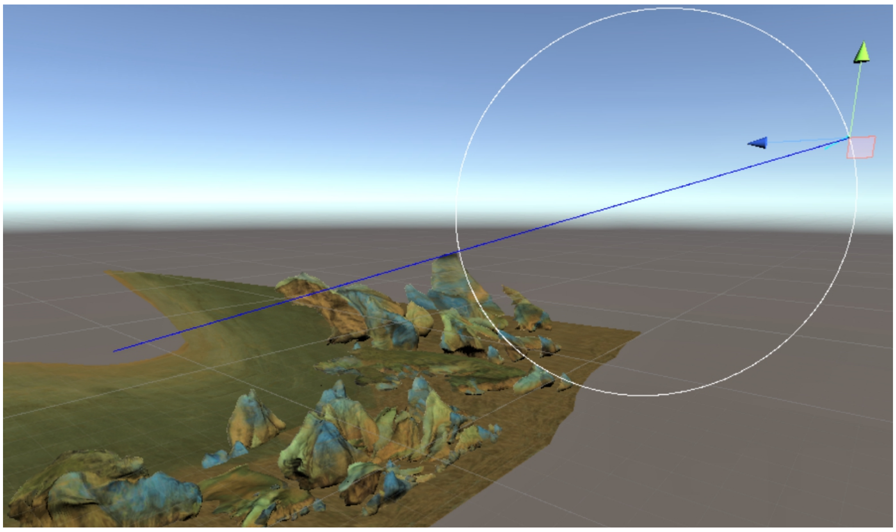

Figure 8.4.2: Synchronize the positions of diegetic viewpoint and the focus point

I added a state-synchronizing script to the viewing platform. This script will set the world coordinates of the viewing platform to the queried position of the diegetic viewpoint, and then changes the x-orientation of the viewing platform to the opposite x-orientation of the queried direction. [Figure 8.4.2]

Last, I loaded three scripts in `void Update()` of the main function of the VR program and set to execute them once per frame. In this way, the VR user can undergo a synchronous diegetic movement when he watches the virtual CHSP. The represented scene that the VR user sees at first glance exactly corresponds to the image in the unrolled frame.

8.4.2 Synchronization between the displayed image in the unrolled frame and the displayed part of the represented scene

The effect of the change of the image in the unfolded frame on the represented scene is reflected in two aspects: the simultaneous change of the displayed part of the ground and the simultaneous change of the displayed objects. These two synchronous associations are accomplished in two steps.

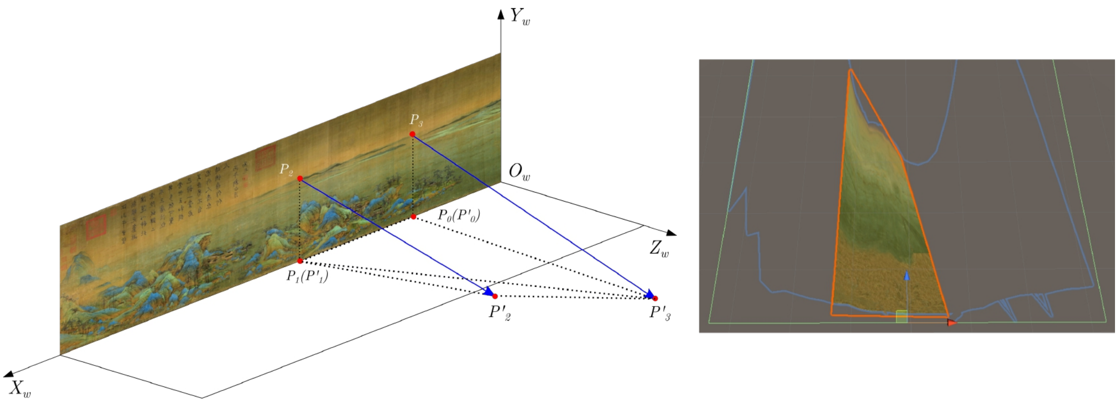

Figure 8.4.3: Create GroundMask

8.4.2.1 Synchronize the image in the unrolled frame and the displayed part of the ground plane of the represented scene

First, I imported the reconstructed ground plane (*ImageryGround*) from 8.3.2 into the virtual environment and added an empty object as its mask (*GroundMask*) to control the display part of the ground plane. I applied the same shader nodes of *ScrollMask* to the shaders of *GroundMask* and *ImageryGround*. The parameters were set as follows.

```
GroundMask: Stencil {
                Ref 1
                Comp Greater
                Pass Replace}
ImageryGround: Stencil {
                Ref 2
                Comp Equal}
```

Second, I added a size-linking script to this *GroundMask* to make its size change correspond to the size of the unrolled frame in real-time. The specific implementation is to create a dynamic *QuadrilateralMesh* as the ground's mask (*GroundMask*), whose four corners correspond to the unrolled frame's four corners. [Figure 8.4.3-left] As all shaped meshes in Unity are composed of triangles, the way to create the *QuadrilateralMesh* is to build two triangles based on four positions. Therefore, the first step of this script is to set up the array of its vertices. Assume the four vertices are P'_0, P'_1, P'_2, and P'_3, their coordinates can be queried from the *Datatable-ground* by using `datatable.Select`. Then, the vertex (*verts*) of the *QuadrilateralMesh* can be set as follows.

```
Vector3[] verts = new Vector3[4]{P'0, P'1, P'2, P'3};
quadrilateral.vertices = verts;
```

The second step of this script is to create two triangle surfaces based on these vertices. To create two triangles, six vertices are required and the vertices of each triangle must be ordered clockwise. Then, the two triangular surfaces can be created by the following codes.

```
int[] tris = new int[6] = new int[6]{
       0, 3, 2,
       0, 2, 1 // 0, 1, 2, 3 are indices of P'0, P'1, P'2, P'3.};
QuadrilateralMesh.trangles = tris;
```

The last step of this script is to add the component of *MeshFilter* to *GroundMask* and change its mesh to the *QuadrilateralMesh* created above.

```
meshFilter = GroundMask.AddComponent<MeshFilter>(); //Add meshFilter component
    to GroundMask.
meshFilter.mesh = QuadrilateralMesh; //give the mesh of the QuadrilateralMesh
    to the GoundMask.
```

Last, I loaded this script into `void update()`. It will generate a dynamic *GroundMask* in real-time based on the four corners of the unrolled frame, as shown in [Figure 8.4.3-right]. In this way, the VR user can control the display area of the ground plane by resizing the unrolled frame.

8.4.2.2 Synchronize the image in the unrolled frame and the displayed objects in the represented scene

The method to achieve this goal is to use the method of `collider.bounds.Contains` to remove the objects that are not in the displayed image in the unrolled frame. This method was implemented in two steps.

(1). Import restored 3D objects

First, I imported all objects restored in 8.3.2 into the virtual environment and added a box collider (*ObjectCollider*) to each object, so that the `collider.bounds.Contains` can detect their existences. Second, I adjusted the *ObjectCollider's* size to be approximately identical to the object's.[19] [Figure 8.4.4-left]

(2). Create an object filter to eliminate objects that do not appear in the unrolled frame.

First, I added an empty object in the virtual environment as the object filter and gave it a mesh collider (*SceneCollider*) component. To accurately separate the inside and outside of the represented scene, this object filter(*SceneCollider*) has to be a *HexahedronMesh* with *GroundMask* as the bottom and top surfaces and has the height, of which h[20] should be higher than the highest object model.

Second, I coded a filter-adjusting script[21] to adjust the shape of the *SceneCollider* according to the shape of *GroundMask* in real-time. As mentioned before, Unity processes and displays all shapes of meshes in triangles. Thus, the method of constructing a *HexahedronMesh* is still to create corresponding triangles. To obtain this *HexahedronMesh*, its vertices should be defined in the first step of this script. According to the four vertices of *GroundMask* and the height h, the eight vertices (*verts*) of the hexahedron can be defined as shown in [Figure 9.4.4-middle], where P'_0, P'_1, P'_2, and P'_3 are the four vertices of *GroundMask*. Their coordinates have been obtained in the making of *GroundMask*. And P'_4, P'_5, P'_6, and P'_7 are the other vertices of the hexahedron. They have the same x-coordinates and z-coordinates as P'_0, P'_1, P'_2, and P'_3, and their y-coordinates is h. Then, the vertex (*verts*) of *HexahedronMesh* can be set as follows.

```
Vector3[] verts = new Vector3[8]{P'0, P'1, P'2, P'3, P'4, P'5, P'6, P'7};
HexahedronMesh.vertices = verts;
```

[19] Theoretically, the mesh collider should be used to set its shape as the same as the object. But the object model imported is not only the 3D object but has a sculptured plane with full-sized CHSP as shown in [Figure 8.3.7]. Therefore, it is needed to create a collider and manually adjust its size to the 3D object's size.]

[20] In the test, we set h to 100.

[21] The code is inspired by http://ilkinulas.github.io/development/unity/2016/04/30/cube-mesh-in-unity3d.html

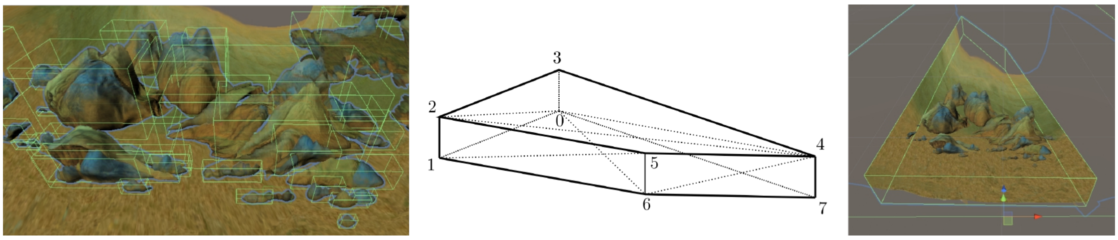

Figure 8.4.4: Create a filter to eliminate the objects outside the represented scene

The second of this script is to make all the surfaces of the triangles that make up this *HexahedronMesh*. As shown in [Figure 8.4.4-middle], a hexahedron is made up of 12 triangles. And the vertices of each triangle need to be arranged clockwise. The code for the triangles array is written below, of which number 0-7 are the indices of the 8 vertices.

```
int[] tris = {
       0, 2, 1,
       0, 3, 2,
       2, 3, 4,
       2, 4, 5,
       1, 2, 5,
       1, 5, 6,
       0, 7, 4,
       0, 4, 3,
       5, 4, 7,
       5, 7, 6,
       0, 6, 7,
       0, 1, 6};
mesh.triangles = tris;
HexahedronMesh.vertices = verts;
```

Then, update the mesh of the *SceneCollider* to the newly created *HexahedronMesh* by,

```
HexahedronMesh.RecalculateBounds();
SceneCollider.shareMesh= HexahedronMesh;
```

Last, I removed the ones that are not in the represented scene. I realized it by using the method of `collider.bounds.Contains`. I iterated through all objects (*ObjectCollider*) to determine if it was in *SceneCollider*.

```
if (SceneCollider.bounds.Contains(ObjectCollider.position) == false){
Object.SetActive(false)}; //if the object is not in the SceneCollider, remove
    it;
else{
Object.SetActive(true)}; //if the object is in the SceneCollider, keep it;
```

In this way, the objects outside the currently represented scene will be removed so that the remaining objects correspond exactly to the objects that appear in the unrolled frame. I loaded this script in `void Update()` to adjust the appearing objects in real-time as the image in the unrolled frame changes.

8.4.3 Synchronization between the position of the diegetic viewpoint and the scene outside the represented scene

According to the construction method of the diegetic scene in 7.4.2, the diegetic scene is a sphere randomly generated centered on the viewer.Then, when the diegetic viewpoint moves, this sphere should move synchronously. At the same time, the diegetic scene is uncertain and it would be differently viewed from different positions. Then, as the position of the diegetic scene changes, its form and content should also change synchronously.

To realize these synchronizations between the position of the diegetic viewpoint and the scene outside the represented scene, I designed the following three steps.

8.4.3.1 Build the ground outside the represented scene

As assumed in 7.4.2, the ground of a sphere-shaped scene of the diegetic world is a roundness centered on the vertical projection of the diegetic viewpoint. The ground outside the represented scene is the rest of the round ground after removing the displayed part of the reconstructed ground in 8.3.2. Therefore, to build this part of the ground, the entire round ground needs to be built. Also, the diegetic viewpoint moves according to the viewer's focus point, the ground is a round that changes dynamically. Therefore the round ground needs to be created in real-time. Here I achieved this goal in three steps.

(1). Create initial ground of diegetic scene

I firstly created an empty object in a virtual environment as the initial ground of the diegetic scene (*RealmGround*) and gave it a component of Mesh Filter and a material with the segmented ground image as its texture. Secondly, I set the shader of the *RealmGround's* material as follows, so that the reconstructed represented ground (*ImageryGround*) can be removed from the *RealmGround*.

```
ground_realm: Stencil{
              Ref 1
              Comp Equal}
```

(2). Code a script to control the center and radius of the round ground

The first step of this script is to get the center and radius of the round. Here, I used `datatable.Select` to query the position of the diegetic viewpoint from *Datatable-viewpoint* that corresponds to the current position of the VR user's focus point on the virtual CHSP. The vertical projection of this position on the ground is the center ($C = (x_c, y_c)$) of the *RoundMesh*. At the same time, I applied the same method to query the 3D position of the projection point from *Datatable-ground*. This point corresponds to the point located on the upper edge of the ground image and has the same x-coordinate as the focus point. According to the diagram in [Figure 7.4.6], this projection point is located on the ground's farthest edge. The distance from this position to the *RoundMesh*'s center is the radius (r) of the ground.

The second step of this script is to create the *RoundMesh* based on the center and radius. The specific method is to create circling isosceles triangles with the *RoundMesh's* center as their apexes and the radius as their two legs, as shown in [Figure 8.4.5-left]. For this, I firstly defined the positions of all triangles' vertices. Assuming that the whole *RoundMesh* is composed of n[22] triangles, then the radian that each triangle corresponds to is $\theta = 2\pi/n$. For an arbitrary point $P = (x_p, y_p)$ on the edge of the *RoundMesh*, the radian of the arc it corresponds to is $\alpha = \theta \times i$,

[22] The smoothness of the round edge depends on the number of triangles. This number can be adjusted according to the fineness of the *ReamlGround* needed.

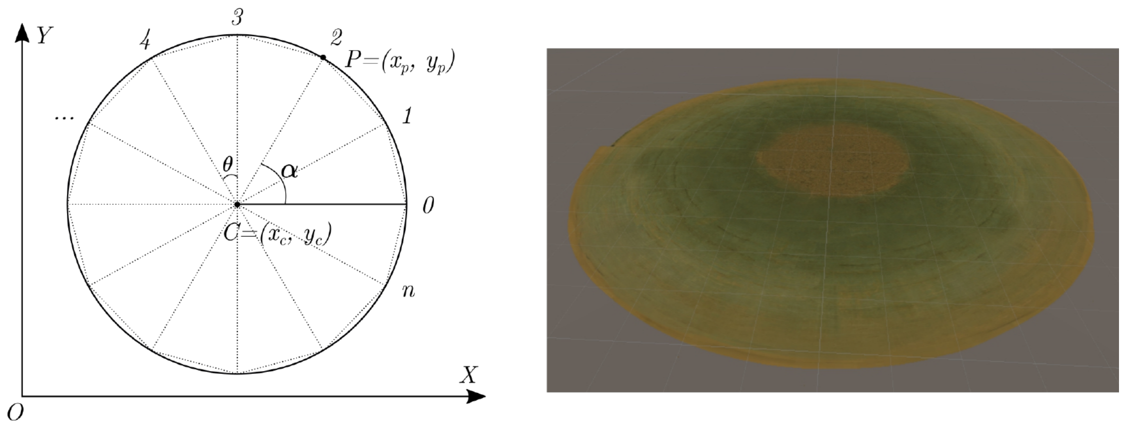

Figure 8.4.5: Create the ground of the diegetic world

where i is the index of P in the sequence starting counterclockwise from the x-direction. And then, the coordinate of P will be, $P = (x_p = x_c + rcos\alpha, y_p = y_c + rsin\alpha)$. [Figure 8.4.5-left] By iterating through their indices, the coordinates of all vertices on the edge of the round can be obtained. The realization codes are as follows.

```
Vector3[] verts = new Vector3[n + 1];
vertices[0] = C;
float deltaTheta = Mathf.Deg2Rad * 360f / segment;
float alpha = 0;
for (int i = 1; i < verts.Length; i++){
        float cosAlpha = Mathf.Cos(alpha);
        float sinAlpha = Mathf.Sin(alpha);
        vertices[i] = new Vector3(cosAlpha * r + C.x, sinAlpha * r + C.y, 0);
        alpha += deltaTheta;};
RoundMesh.vertices = verts;
```

Secondly, I connected three vertices to form the triangles circling the *RoundMesh's* center in the clockwise order as follows. [Figure 8.4.5-left]

```
0,2,1;
0,2,3;
0,3,4;
...
0, n, 1
```

And this connection was realized by the following codes.

```
int[] tris = new int[n * 3];
for (int i = 0, j = 1; i < n * 3 - 3; i += 3, j++){
        triangles[i] = 0;
        triangles[i + 1] = j + 1;
        triangles[i + 2] = j;};
triangles[n * 3 - 3] = 0;
triangles[n * 3 - 2] = 1;
triangles[n * 3 - 1] = n;
RoundMesh.triangles = tris;
```

Thirdly, I updated the RealmGround's mesh to the RoundMesh by,

```
meshFilter = ReamlGround.getComponent<MeshFilter>();
meshFilter.mesh = RoundMesh;
```

Last, I loaded this script in the `void Update()` of the main function of the VR program so that the *RealmGround* will be generated simultaneously as the diegetic viewpoint moves. [Figure 8.4.5-right]

8.4.3.2 Build the sky

According to the construction method of the diegetic scene in 7.4.2, the diegetic sky is a sphere expanding from the viewer to the end of the ground.Therefore, the sky is a dynamic sphere that would change synchronously as the position of the diegetic viewpoint and the end of the ground change. Here, I conducted two steps to build this dynamic sky.

First, I created a sphere in the virtual environment as the sky (*SkySphere*)[23]. Since the scene of the diegetic world is centered on the diegetic viewpoint, I placed *SkySphere* in *VR camera* as its child object, so that its center will follow the movement of *VR camera*.

Second, I rendered the sky image on the inner surface of this sphere. To realize it, I gave created a sky material with the segmented sky image as its texture and applied the following shader.[24]

```
Shader "Custom/Sky" {
   Properties{_MainTex("Base (RGB)", 2D) = "white" {}}

      SubShader{
         Tags { "RenderType" = "Opaque" }
         Cull front
         LOD 100

         Pass {CGPROGRAM
                 #pragma vertex vert
                 #pragma fragment frag

                 #include "UnityCG.cginc"

                 struct appdata_t {
                     float4 vertex : POSITION;
                     float2 texcoord : TEXCOORD0;};

                 struct v2f {
                     float4 vertex : SV_POSITION;
                     half2 texcoord : TEXCOORD0;};

                 sampler2D _MainTex;
                 float4 _MainTex_ST;

                 v2f vert(appdata_t v){
```

[23]The initial radius needs to be set according to the size of the reconstructed represented scene. In the test, I set it as 800.

[24]The codes were inspired by Bernie Roehl's method. http://bernieroehl.com/360stereoinunity/

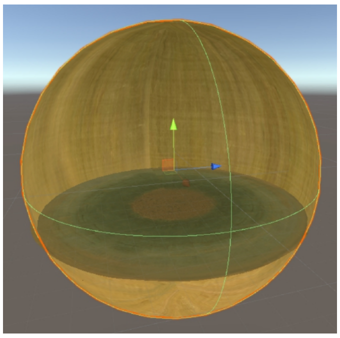

Figure 8.4.6: Create the sky of the diegetic world

```
                v2f o;
                o.vertex = UnityObjectToClipPos(v.vertex);
                v.texcoord.x = 1 - v.texcoord.x;
                o.texcoord = TRANSFORM_TEX(v.texcoord, _MainTex);
                return o;}

            fixed4 frag(v2f i) : SV_Target{
                fixed4 col = tex2D(_MainTex, i.texcoord);
                return col;}
    ENDCG}}}
```

Third, I added a sky-adjusting script to adjust *SkySphere's* radius r in real-time. According to the diagram in [Figure 7.4.6], the sphere's radius is the distance between the diegetic viewpoint and the farthest point on the ground of the diegetic world. Then its radius (r) can be obtained by, `r = Vector3.Distance (the position of the sphere's center, the position of the projected point on the further edge of` *`ImageryGround`*`)`. The position of this projection point has been obtained in 8.2.3. Then, I updated the radius of the SkySphere by `SkySphere.transform.scale = new Vector3(r, r, r)`. [Figure 8.4.6]

Last, I loaded this script in `void Update()` of the VR program to adjust the radius of the sky in real-time.

8.4.3.3 Create objects outside the represented scene

As explained in 7.4.2, I plan to use the method of "random generation" to create the uncertain scene outside the represented scene. Briefly, randomly replicating the objects from the represented scene and arranging them in the scene outside the represented scene. This method was implemented in two steps.

First, I created an empty object (*ObjectEmpty*) at the position of each imported object. I adjusted their z-orientations to the same directions as the projection rays passing through their ground-contact points. [Figure 8.4.7-left] These directions were retrieved from the *Datatable-vector*. And then, I placed each object in its corresponding *ObjectEmpty* as a child object.

Second, I added an object-cloning script to randomly replicate and arrange objects in the scene outside the represented scene. In this script, I used a "for" loop and set the desired amount of cloned objects as the length of the counter of this loop.[25] In each loop, I firstly randomly selected the index of the *ObjectEmpty* by using `UnityEngine.Random.Range(0, the total`

[25]The length of the counter can be adjusted according to the density of the objects needed.

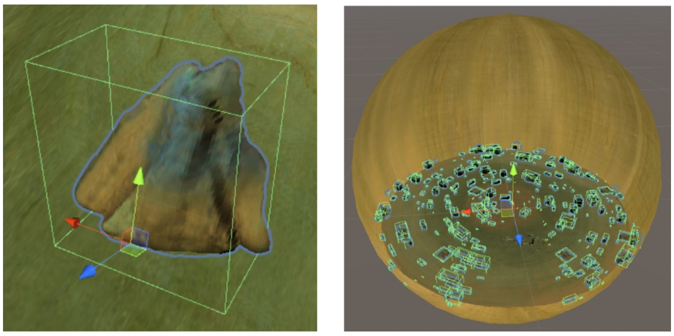

Figure 8.4.7: Clone and arrange objects in the scene outside the represented scene

`amount of ObjectEmptys)`. And then, I got the selected object by `ObjectEmptys.transform.GetChild(index).gameObject`. Secondly, I used `Random.insideUnitCircle*` to randomly generate a position (*RandomPosition*) within the round ground, where r is the radius of the Realm-Ground. Thirdly, I cloned the selected *ObjectEmpty* with its child object as a *ClonedObject* and placed it on the *RandomPosition* by using `Object.Instantiate(the selected ObjectEmpty, RandomPosition)`. Lastly, I rotated the cloned *ObjectEmpty* to make its z-orientation towards the *VR camera* by `ClonedObject.transform.Lookat(VR Camera.transform)`, so that all *ClonedObject* would face the VR user all the time.

Since the cloned objects were arranged randomly, some of them may be placed in the represented scene. Therefore, after finishing for loop, I need to remove those misplaced cloned objects. For this, I iterated through all cloned objects and determined whether the cloned objects are located in the *SceneCollider*[26] by using `SceneCollider.bounds.Contains(ClonedObject.position)`. If the the cloned object is in the *SceneCollider*, it will be eliminated by `Destroy(the misplaced ClonedObject)`.

Last, I loaded this script in `void update()`. But the scene outside the represented scene should not change per frame, but should change only the position of the diegetic viewpoint has changed. Therefore, I added a condition judgment to determine the case to execute the object-cloning. The script shows as follows.

```
if(currentVRCamera.position = ! previousVRCamera.position)
Destroy(all cloned objects); // clear objects in the scene outside the
    represented scene;
call object-cloning script;
```

Finally, I integrated all data, models, and scripts and compile them as a VR application. The images in [Figure 8.4.8] show the operation of the VR application.

Summary

This chapter elaborated on the program implementation for synchronously simulating the viewing experience and the diegetic experience of "touring in the CHSP" on the VR platform. This program is realized in three steps.

[26] *SceneCollider* is the *HexahedronMesh* created in 8.4.2.

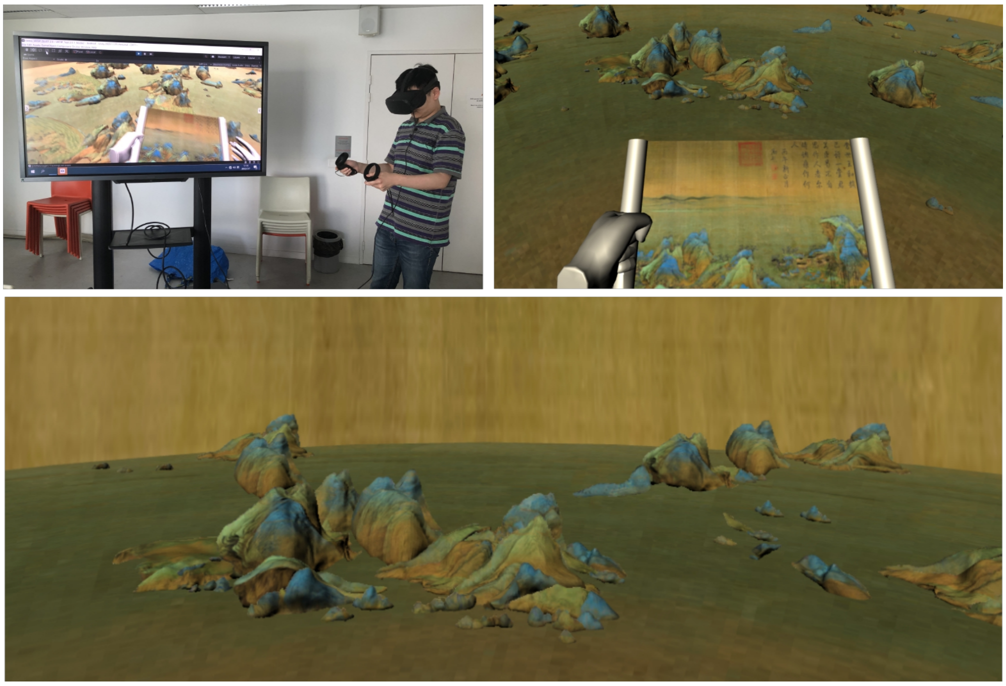

Figure 8.4.8: Operation of the virtual exhibition system

Step 1: Data preparation

In this step, three sets of data were computed: the directions of the projection rays passing through all pixels on the CHSP, the positions of the projected points in 3D space corresponding to all pixels on the CHSP, and the positions of the viewer's viewpoint in the scene of the diegetic world corresponding to all pixels on the CHSP. These data are computed according to the computing methods developed in 7.2, 7.3, and 7.4. All results of computations were output in the form of datatable.

Step 2: Model preparation

In this step, two scenes were constructed: the viewing scene, including a viewing platform and an interactive CHSP; the represented scene of the entire CHSP. The former was achieved by the ordinary 3D modeling method by referring to a house in the CHSP. The latter was completed by the methods developed in 7.4.

Step 3: Synchronization program development

This program aims to realize the following three synchronous correspondences.

(1). Correspondence between the VR user's focus point on the virtual CHSP and the viewing platform's position and orientation

This correspondence was achieved in three steps. First, get the intersection of the virtual CHSP and VR user's sightline (ray) by RaycastHit. Second, query the position and direction of the diegetic viewpoint corresponding to this intersection. Third, translate the position and orientation of the viewing platform.

(2). Correspondence between the image in the unrolled frame and the displayed part of the reconstructed represented scene

This correspondence was implemented in two steps. First, create a ground mask to control the displayed part of the reconstructed ground and correlate the four corners of the unrolled frame and the four corners of the mask. Second, create a collider to filter the objects of the display image in the unrolled frame.

(3). Correspondence between the position of the diegetic viewpoint and the scene outside the represented scene

This correspondence was implemented in three steps. First, build the round ground based on the center and radius calculated by the method in 7.4.2.1. Second, create the spherical sky based on the center and radius calculated by the method in 7.4.2.1. Third, randomly clone and arrange restored objects in the represented scene and remove the cloned objects that appear in the recreated scene by collider filter.

The synchronization program was compiled into a VR application that runs on the Oculus Quest.

Chapter 9

Discussion

This chapter will discuss the result of this research. First, the test and evaluation of this virtual exhibition system will be presented. Second, the main contributions and significance of this research will be discussed. Third, the limitation of this study will be addressed.

9.1 Tests of virtual exhibition system

This virtual exhibition system was developed based on theoretical inference and geometrical analysis. Its validity needs to be tested by subjective experience.

For this reason, I invited 15 participants to test the virtual exhibition system. Of these, 10 of them did not know the CHSP before and 6 participants had experience in operating VR devices. The participants were required to conduct four tests, which correspond to four results achieved.

9.1.1 Test 1: Emulation of virtual CHSP

The test was designed to evaluate how well the virtual CHSP emulated the actual CHSP in terms of appearance and manipulation. This test was conducted in three steps. First, the participant was asked to watch a demonstration of the real CHSP and its manipulation in reality. Second, the participant was instructed to manipulate the virtual CHSP in the virtual operating system. Third, the participant was required to test the manipulation of the virtual CHSP, such as unrolling, rolling up, stretching, etc. Fourth, the participant was asked to answer two questions after testing.

- *How easy is it to control the virtual CHSP?*

- *To what extent does the virtual CHSP emulate the real CHSP? in terms of appearance and interaction.*

The results show that, 67% of the participants found the virtual CHSP difficult to operate. However, after the second demonstration, this number dropped to 16.7%. All participants thought that the virtual scroll simulated 90% of the real CHSP's appearance. But in terms of manipulation, the virtual CHSP only reached 60%. They thought that the virtual CHSP only simulated the basic manipulations of unrolling and rolling up in one direction one time, but not in both directions at the same time.

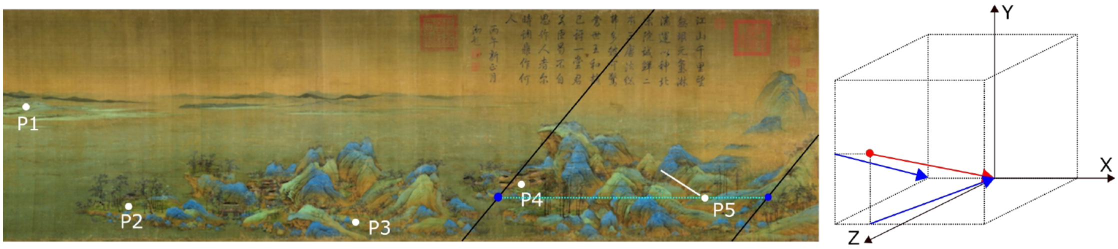

Figure 9.1.1: Estimate projection vectors

9.1.2 Test 2: Computation of the diegetic viewpoint's orientation and position

The purpose of this test is to verify whether the computed diegetic viewpoint position and orientation are consistent with human perception.

According to the figure in [Figure 5.2.2], the estimations of the orientation and position of the diegetic viewpoint were based on three geometrical relations. First, the diegetic sightline has the opposite direction of the projection ray. Second, the diegetic viewpoint locates on the projection ray. Third, the height of the small projection plane (the CHSP) is equal to the height of the viewer's field of view.

Of those, the direction of the diegetic viewpoint can be judged directly from the tilt of the object image. But it is more difficult to determine the position directly because it requires participants to combine the last two conditions simultaneously. Then an indirect approach is needed. There is a noteworthy phenomenon that, when the conditions are satisfied simultaneously, the object image on the small projection plane will have the same dimensions as the corresponding object in vision. Also according to the formation of cell-scene image, the latter is related to the distance between the diegetic viewpoint and the object. Then if the size of the restored object seen by the participant in the virtual scene is the same as the size of the object image he sees in the virtual CHSP, then the position of the participant to the restored object at this time is equal to the distance between the diegetic viewpoint and the object in the participant's perception. In this case, if the orientation of the diegetic viewpoint can also be satisfied to match the participant's perception, then it can be determined that its position matches the participant's perception.

Based on this inference, I designed the following test method.

First, the participant was required to estimate the directions of the projection rays passing through five selected points in [Figure 9.1.1-left] and then respectively draw the vertical and horizontal component vectors (blue arrows) of each projection vector (red arrow), as shown in [Figure 9.1.1-right].Second, the participant entered the virtual environment. Therefore, he was asked to move to these five diegetic positions and then compared the sizes of the objects in front of him and the object images in the unrolled frame. Third, I calculated the unit vectors of all the projection rays estimated by the participants and illustrated them in five spheres shown in [Figure 9.1.2] The blue dots are those estimated by participants and the red dots in these figures are the unit vectors I calculated.

(1). Result of the estimation of projection rays' directions

It can be seen from the figures, the calculated directions of the projection rays passing through $P1$, $P2$, $P3$, and $P4$ are relatively close to the average directions estimated by the participants. While for $P5$, the difference is relatively larger. The reason caused this deviation is that the

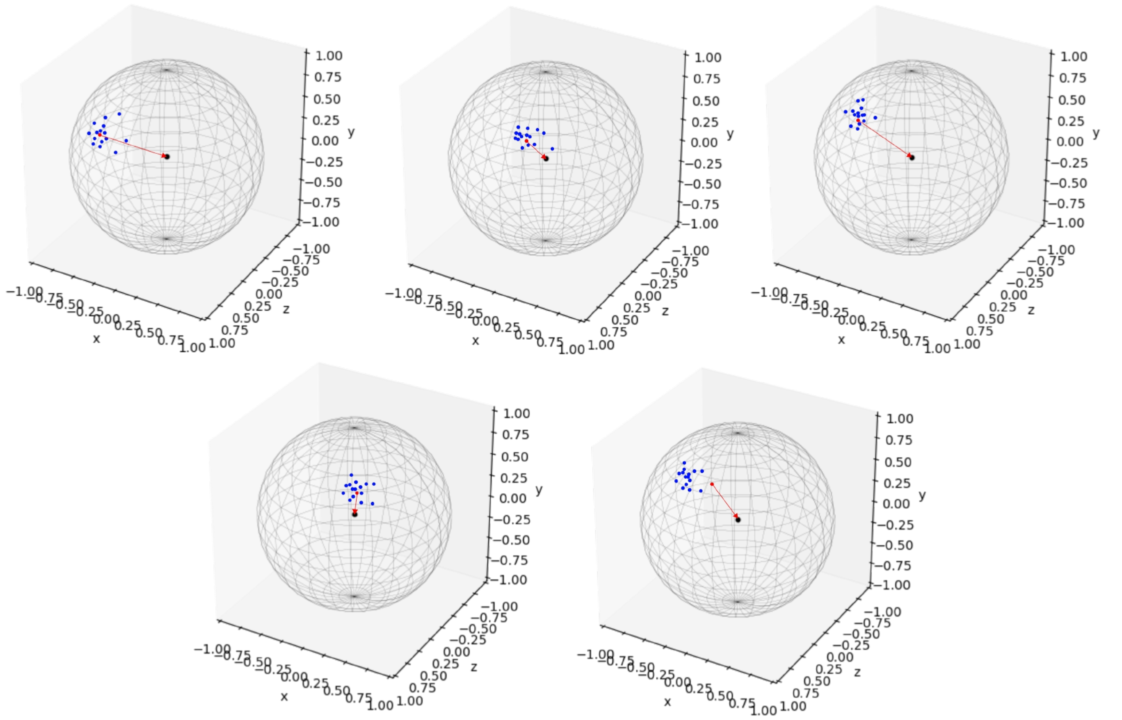

Figure 9.1.2: Testing result of the estimation of projection ray's direction

reference diagonal I selected for calculation is not the same as the diagonal the participants referenced. The one I selected is the right edge of one mountain shown by the black diagonal in [Figure 9.1.1], while the participants selected the left edge of the side shoulder of the other mountain, as shown by the white diagonal.These diagonals have different inclinations, which imply different directions. Since $P5$ is very close to the white diagonal, participants estimated the projection direction directly according to this diagonal. And I obtained it by calculating the *SLerp* of the directions of the projection rays passing through the reference points on the left and right sides of $P5$.[1] Therefore, the different reference diagonals lead to different results.

(2). Result of the comparison of objects' sizes

The histogram in [Figure 9.1.3] illustrates the result of this test. From this histogram, it can be concluded that most participants believed the objects seen from the diegetic positions corresponding to $P2$, $P3$, and $P4$ had approximately the same sizes as their images in the unrolled frame. While the dimensions of the objects seen from the diegetic positions corresponding to $P1$ and $P5$ differed considerably from their images. According to the feedback, the restored objects looked smaller than their images from the diegetic position corresponding to $P1$, but bigger from the diegetic position corresponding to $P5$.

The deviation corresponding to $P1$ may relate to the estimated size reduction ratios for restoring the object's relative size in 8.2.3.3. This scale was extrapolated from two near houses, which originally had a deviation. And this deviation was amplified when multiplying far distances. The deviation corresponding to $P5$ may be due to the different selections of diagonals.

[1] See the diagonals selection is shown [Figure 8.2.1].

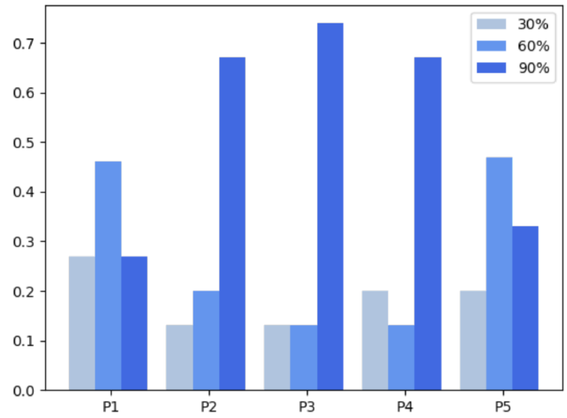

Figure 9.1.3: Result of scale estimations

(3). Result of the estimation of diegetic positions

As previously inferred, if the condition that the computed direction of the diegetic viewpoint is consistent with the one the participant estimated and the condition that the restored object's size is consistent with what the participant sees in the virtual space can be satisfied at the same time, then the computed position of the diegetic viewpoint is correct. Based on this inference, I combined the two results. The result shows that the calculated positions of the diegetic viewpoints corresponding to $P2$, $P3$, and $P4$ are close to that of the participant perceived, but that of the diegetic viewpoints corresponding to $P1$ and $P5$ are not.

In summary, the calculated directions and positions are generally consistent with the participant's perception. This indicates that the calculation method I used is correct. However, there are still deviations in individual cases. These deviations were caused by inaccurate references, which means the I need to provide more refined reference data.

9.1.3 Test 3: Evaluate the construction of the diegetic scene

The diegetic scene contains two parts: the reconstructed scene is represented in the CHSP. and the scene outside the represented scene. Therefore, this test was divided into two parts as follows.

(1). Evaluate the reconstruction of the represented scene

This evaluation focused on two items: the depth data corresponding to all the points on the ground image, the correspondence between the displayed part of the reconstructed scene and the image in the unrolled frame during the VR interaction.

First, the participant was asked to estimate the 3D positions of the five selected points on the ground and mark the estimations on a top-view plan. The figure in [Figure 9.1.4] illustrates the result. The blue dots represent participants' estimations, the black dots are the positions I calculated. As the figure shows, the calculated 3D positions are in general consistent with the participants' estimations in terms of depth. However, in terms of specific location, the calculated 3D positions show differences from participants' estimations.

According to the analysis, these differences come from the references used for computation

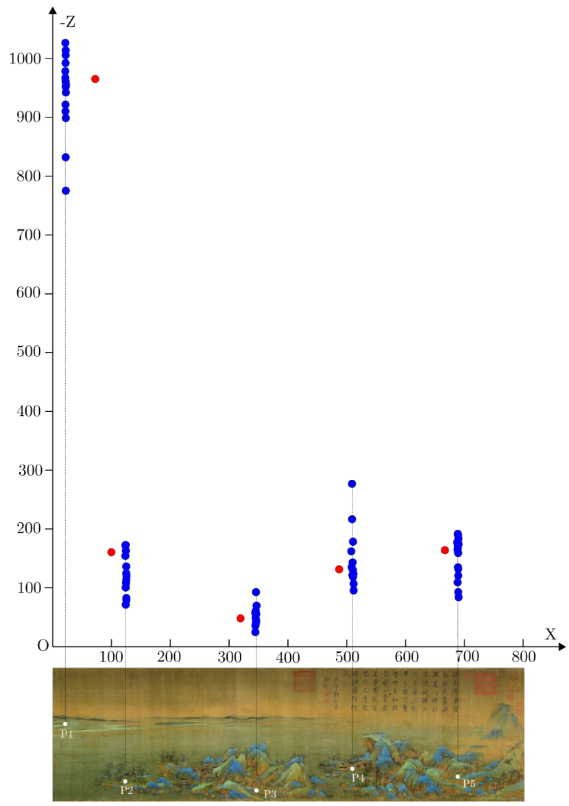

Figure 9.1.4: Result of estimation of the ground points' 3D positions

and estimation. In computation, I used projection ray as the computation basis. The projection ray is 3D, it may have deflection in the horizontal direction. While the participants did not consider this deflection.

Second, the participant was to move to different diegetic positions and compare the reconstructed scene in front of him to the image in the unrolled frame. The result of this test shows that 73% of the participants thought that the consistency between the represented and the image reached 80% or above. While there were 27% of the participants argued that the consistency was lower than 80%.

According to their feedback, the deviation lies in the transition of the represented scene at the boundary. In the current situation, restored objects around the boundary would suddenly appear or disappear all at once, instead of gradually fading in or out like the object images in the virtual CHSP. This phenomenon is particularly evident when a large object is at the boundary. The reason causes this deviation is that, the applied method of `collider.bounds.Contains` for filtering objects only has two options, completely inside the collider or not. Therefore, when an object is located on the boundary of the collider, it will be treated as being outside and will be removed completely. This method is efficient but not refined enough. Therefore, a more refined way needs to be developed.

(2). Evaluate the construction of the scene outside the represented scene

The purpose of this test was to examine the extent to which the randomly created scene matched the participant's imagination.

First, the participant was required to select 5 regions in the CHSP and describe the scenes outside the frame. Second, the participant was asked to move to the corresponding positions of these regions in the virtual environment and compare the randomly created scenes to their initial imaginations.

The result shows that 20% of participants thought that the consistency of the randomly created scene and their imagination reached 50%, while the rest thought this number was below 50%. This result is in line with expectations, as I want to make this scene uncertain.

In addition, this test exposed a problem in the current system. Due to the small variety of objects, the same object may be cloned several times and appear in the same scene, which makes the generated scene very unnatural. Therefore, the range of cloned objects needs to be extended and the number of clones of the same object model needs to be limited.

9.1.4 Test 4: Evaluate the general experience of virtual tour

In this test, the participants was asked to freely travel through the virtual environment and then evaluate their experiences during this process.

- Do you think you were free to move in the virtual space?

- Do you think you were traveling in one space or through multiple spaces?

- Can you briefly describe your experience, from the aspects of the path, degree of freedom, and sensation?

The result shows that, 80% of the participants mentioned they had a joyful experience of "freely traveling in the painting". 67% of participants had the experience of "moving through different scenes", in the condition that they were not told about the CHSP as a representation of multiple scenes.

This result proves that the virtual display system has realized the simulation of "touring in the CHSP, which means it can be used as an interpreter in the future virtual exhibition of the CHSP.

9.2 Discussion of virtual exhibition system

In this research, a virtual presentation system was finally developed. This system will allow the visitor to manually manipulate the CHSP, which solves the contradiction between conservation and handful manipulation requirements. Meanwhile, the system can simulate the diegetic experience of the "tour into the CHSP", which provides an intuitive way to get a complete aesthetic experience of the CHSP. In this sense, this system can provide a solution to implement two shortcomings[2] in current exhibitions of CHSPs.

Compared to current VR-based exhibition systems, this system has the following novelties.

(1). Spatialization based on theoretical inference

The diegetic scene and diegetic movement presented in this system were the results of a spatialization based on ancient aesthetic theory, space theory and painting theory.

Compared with the most commonly used method based on the individual subjective experience of VR makers, this method can avoid the influence of modern and individual thinking, so that the resulting diegetic scene and diegetic movement can be much closer to the diegetic experiences described by the ancient viewers.

(2). Geometrization based on CHSP's visual structure

[2] See two shortages in 1.2

In this system, the geometric transformation model, on which the diegetic viewpoint was calculated and the represented scene was reconstructed, was established based on the visual structure of CHSP.

In establishing this transformation model, layered and period-by-period analysis methods were adopted. These methods bring two advantages to this transformation model.

First, the layered analysis took into account the impact of all painterly operations on the transformation model during CHSP formation, such as projection, graphical composition, etc. This avoids the partial interpretation that comes from simply assigning it to a specific type of projection, thus making the transformation model relatively comprehensive. Second, the analysis by period considered the influences of different compositions of the CHSPs in different periods on the transformation models. In this way, the characteristics and differences of CHSP transformation models in different periods can be derived, so that a specific transformation model can be selected for a particular CHSP.

(3). Algorithm with projection ray as association

In this system, the algorithms used for computation and construction were developed using projection rays as data associations.

Compared to the methods used in some current projects, such as using the angle of the projection plane to the ground (Ma, 2011, 2012) or visual perception (Kadar, 2008) to estimate depth and viewpoint position, this approach takes into account the difference in 3D position caused by the horizontal deviation of the diegetic viewpoint's direction, making the calculated positions of the diegetic viewpoint and the 3D positions of ground points more accurate. Meanwhile, this method can correlate the position of the diegetic viewpoint and the 3D position of the ground point to each other, which makes it possible to realize the synchronous correspondences between the diegetic viewpoint and the focus point on the CHSP.

(4). Synchronized presentation and interpretation

In this virtual exhibition system, the presentation scene (the virtual viewing scene) was nested within the interpretation scene (the diegetic scene of the CHSP), rather than set them up separately as is the case with most current virtual exhibits.

This approach enables VR users to simultaneously experience the corresponding changes between the interaction with the CHSP and the movement in the diegetic scene, which achieve a real-time interpretation of the CHSP. At the same time, this approach puts the control of diegetic movement in the hands of the VR user. In this way, there is no need to preset a fixed diegetic movement route. Throughout the experience, the VR user is completely free, which is more in line with the ancient way of interaction between the ancient viewer and CHSP.[3]

Based on the above features, this system can be used as a complement to the current CHSP exhibition in terms of presentation and interpretation. However, it should be noted that this interpretation this system provides is not a final aesthetic experience of the CHSP. According to Chinese aesthetics (Ye, 2009), the interpretation of the masterpieces is infinite. It is like the process when we watch the works of Picasso and Cézanne, we will not stop at the question of who or what is portrayed but will keep searching for more profound implications, such as the symbolic meaning, emotional expression, etc. Meanwhile, it is a stair-like process. Only if we have reached a stage can we climb up to the next higher one. In this sense, any interpretations will not be a final answer but a stage leading to various deeper comprehensions. For example, the iconological interpretation may direct us to think of the symbolic meaning of paintings (Panofsky, 1991). Gestalt psychological analysis will lead us to discuss the illusion in the painting (Arnheim,

[3]See the direction of watching a CHSP in 3.3.

1954). Architectural theorists' structural explanation will draw our attention to think about the spatial structure of the represented space (Rowe, 1963). In this sense, the interpretation this provides is establishing a common basis for the visitor to think of the classical Chinese experience of "careless" traveling 逍遥[4] in the CHSP.

9.3 Limitation of the virtual exhibition system

The virtual presentation system developed in this study generally achieved the initial purpose. However, it is still a prototype with limitations.

(1). Not fully automated computation and reconstruction process

In the development of the system, I have implemented the algorithm modules in Chapter 7 into corresponding scripts, which would largely reduce the workload of the VR maker in future development. But specific steps still need to be done manually by the VR maker. For example, the segmentation of the object image, which requires the VR maker to complement its obscured parts, and the depth differentiation of object image, which needs the VR maker to create the depth map of the object image by perception. All of these tasks require a huge amount of labor.

And in the current research on computer vision, there are already some relatively mature techniques that can be applied to these processes. Such as thresholding, watershed, edge detection, machine learning, etc. However, most of these methods were developed for photographs, so they need to be adjusted to the characteristics of images in Chinese paintings, such as visual structure, blurred borders, etc. Therefore, how to adapt these technologies and apply them to the development of virtual exhibition systems is a direction for future research.

(2). Unnatural random scene

For the construction of the scene outside the represented scenes, I developed a random generation approach. This approach randomly selects and duplicates the restored objects to build an uncertain scene.

Although this method can meet the basic requirements, there would be many duplicate objects in the scene due to the limitation of the number of restored objects that can be cloned.

A simple way to solve this problem is to expand the range of objects to be cloned. This method is feasible for the current CHSP of *A Thousand Miles of Rivers and Mountains*, as it is large and contains many objects. However, for other smaller CHSPs containing fewer objects, such as the painting of *Autumn Colors on the Que and Hua Mountains*, a more refined random generation method needs to be developed. And using Artificial Intelligence to generate this scene can be an option. Future research will also work in this direction.

(3). Single perceptual experience

Although VR handles were used to manipulate the virtual CHSP, the simulation of the diegetic experience of "touring in the CHSP" in the current virtual exhibition system is mainly a visual experience. Other perceptual experiences were not taken into account, such as sound. As the 3D positions of all objects can be obtained, then in future development, the corresponding sounds of objects and environment can be added to increase the immersion of the diegetic scene.

(4). Missing social function

There are two exhibition scenes of CHSP that were commonly mentioned in the literature or represented in paintings. One is the viewer watched the painting alone in his study, which

[4] See the definition of "careless" traveling in 2.2.3.

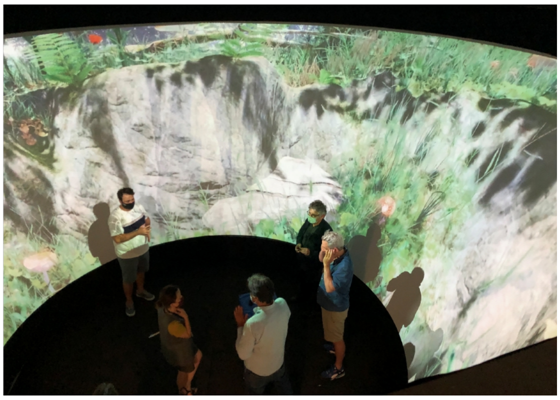

Figure 9.3.1: Viewing scene in the painting of *Elegant Reunion in the Western Garden*; Panoramic presentation in CORAULIS

has been simulated in the current exhibition system. The other is in a gathering of literati, they watch a CHSP together. This scene can be seen in Liu Songnian's 刘松年(1174 - 1224) *Elegant Reunion in the Western Garden*西园雅集图, which illustrates a literati's party in the Song Dynasty. [Figure 9.3.1-right] As the image illustrates, one literati manipulates the CHSP, while the others watch it along with his scrolling action. If we see the aesthetic experience of one viewer watching the CHSP as his single touring in the diegetic scene, then this scene can be seen as a viewer leading his guests to travel in a diegetic world. This case should also be simulated in the virtual exhibition system of CHSP.

However, the current program was developed for a head-mounted VR device, which allows only one person to participate at a time. While the rest of the audience can only share the VR user's view through an external screen. [Figure 8.4.8] Therefore, to simulate the above group viewing scenario, a virtual showroom that can accommodate a larger audience is necessary. Such as, the CORAULIS[5] with a 360 screen. [Figure 9.3.1-right] In future research, this virtual exhibition system will be deployed in this virtual room.

[5]Centre d'Observation en Réalité Augmentée et Lieu d'Immersion Sonore in ENSA Nantes. https://www.nantes.archi.fr/lensa-nantes/coraulis/

Bibliography

[1] Anjyo, Ken. Tour into the picture as a non-photorealistic animation. *ACM SIGGRAPH Computer Graphics*, 33(1):54–55, 1999.

[2] Arnheim, Rudolf. *Art and visual perception: A psychology of the creative eye*. Univ of California Press, 1954.

[3] Arnheim, Rudolf. Perception of perspective pictorial space from different viewing points. *Leonardo*, 10(4):283–288, 1977. Publisher: The MIT Press.

[4] Arnheim, Rudolf. *Visual thinking*. Univ of California Press, 1997.

[5] Aubry, Mathieu, Russell, Bryan C., and Sivic, Josef. Where was this picture painted?-Localizing paintings by alignment to 3D models. 2014.

[6] Bachhofer, Ludwig. The Representation of Space in Chinese Painting during the First Thousand Years of Chiristian Era. 1931.

[7] Bachhofer, Ludwig. Chinese Landscape Painting in the Eighth Century. *The Burlington Magazine for Connoisseurs*, 67(392):189–191, 1935. Publisher: JSTOR.

[8] Barbieri, Loris, Bruno, Fabio, and Muzzupappa, Maurizio. Virtual museum system evaluation through user studies. *Journal of Cultural Heritage*, 26:101–108, 2017. Publisher: Elsevier.

[9] Belting, Hans. *Florence and Baghdad: Renaissance art and Arab science*. Belknap Press of Harvard University Press, 2011.

[10] Benjamin, Walter. *The work of art in the age of mechanical reproduction*. Penguin UK, 2008.

[11] Berque, Augustin. *Thinking through landscape*. Routledge, 2013.

[12] Bolewski, Christin. 'Shan-Shui-Hua'-traditional Chinese landscape painting reinterpreted as moving digital visualisation. In *EVA*, 2008.

[13] Booker, Peter Jeffrey. A history of engineering drawing. 1963.

[14] Burke, Verity, Jørgensen, Dolly, and Jørgensen, Finn Arne. Museums at home: Digital initiatives in response to COVID-19. *Norsk museumstidsskrift*, 6(2):117–123, 2020.

[15] Cahill, James. Chinese painting. 1960. Publisher: Skira.

[16] Cahill, James. *Hills beyond a river: Chinese painting of the Yuan Dynasty, 1279-1368*. Weatherhill, 1976.

[17] Cahill, James. *Parting at the shore: Chinese painting of the early and middle Ming dynasty, 1368-1580*. Weatherhill, 1978.

[18] Cantor, Lea. Zhuangzi on 'happy fish'and the limits of human knowledge. *British Journal for the History of Philosophy*, 28(2):216–230, 2020. Publisher: Taylor & Francis.

[19] Carrozzino, Marcello and Bergamasco, Massimo. Beyond virtual museums: Experiencing immersive virtual reality in real museums. *Journal of Cultural Heritage*, 11(4):452–458, 2010.

[20] Carrozzino, Marcello, Evangelista, Chiara, Brondi, Raffaello, Tecchia, Franco, and Bergamasco, Massimo. Virtual reconstruction of paintings as a tool for research and learning. *Journal of Cultural Heritage*, 15(3):308–312, 2014.

[21] Chen, Baozhen 陈葆真. 洛神赋图与中国古代故事画. 石头出版股份有限公司, 2011.

[22] Chen, Chuanxi 陈传席. 宗炳《画山水序》研究. 美术大观, (1):37–41, 2016.

[23] Chen, Shouxiang 陈绶祥. 近大远小以大观小变时变空——中国画透视刍议. 美术, 7, 1982.

[24] Chu, NS-H and Tai, Chiew-Lan. Animating Chinese landscape paintings and panorama using multi-perspective modeling. In *Computer Graphics International 2001. Proceedings*, pages 107–112. IEEE, 2001.

[25] Clunas, Craig. *Superfluous things: material culture and social status in early modern China*. University of Hawaii Press, 1991.

[26] Clunas, Craig. *Art in China*, volume 2. Oxford University Press, USA, 1997.

[27] Clunas, Craig. *Pictures and visuality in early modern China*. Reaktion Books, 1997.

[28] Clunas, Craig. *Empire of Great Brightness: Visual and Material Cultures of Ming China*. Reaktion Books, 2007.

[29] Clunas, Craig. *Chinese painting and its audiences*. Princeton University Press, 2017.

[30] Costall, Alan. Beyond linear perspective: A cubist manifesto for visual science. *Image and Vision Computing*, 11(6):334–341, 1993. Publisher: Elsevier.

[31] Criminisi, Antonio, Kemp, Martin, and Zisserman, Andrew. Bringing pictorial space to life: computer techniques for the analysis of paintings. *Digital art history: A subject in transition*, 1:5, 2005. Publisher: Intellect Books.

[32] Dufrenne, Mikel. *The phenomenology of aesthetic experience*. Northwestern University Press, 1973.

[33] Fan, ZhenBao, Zheng, XianJun Sam, and Zhang, Kang. Computational analysis and eye movement experiments of white space in Chinese paintings. In *2015 IEEE international conference on progress in informatics and computing (PIC)*, pages 301–306. IEEE, 2015.

[34] Feng, Wei. *An Examination of Chinese Pre-modern Visual Media, Its Influence on Landscape Ideology , Aesthetics and Relationship to Landscape Experience*. PhD thesis, University of Sheffield, 2005.

[35] Feng, Youlan 冯友兰. *A short history of Chinese philosophy*, volume 91098. Simon and Schuster, 1948.

[36] Fong, Wen. *Summer Mountains: The Timeless Landscape*. Metropolitan Museum of Art, 1975.

[37] Fong, Wen. *Beyond Representation: Chinese Painting and Calligraphy, 8th-14th Century*, volume 48. Metropolitan Museum of Art, 1992.

[38] Fong, Wen C. Toward a structural analysis of Chinese landscape painting. *Art Journal*, 28(4):388–397, 1969.

[39] Fong, Wen C. Images of the Mind. *Princeton: The Art Museum, Princeton University*, 1984.

[40] Fong, Wen C. Why Chinese painting is history. *The Art Bulletin*, 85(2):258–280, 2003.

[41] Freedberg, David and Gallese, Vittorio. Motion, emotion and empathy in esthetic experience. *Trends in cognitive sciences*, 11(5):197–203, 2007. Publisher: Elsevier.

[42] Fu, Xinian 傅熹年. 中国古代建筑史. 中国建筑工业出版社, 北京, 2001.

[43] Furferi, Rocco, Governi, Lapo, Volpe, Yary, Puggelli, Luca, Vanni, Niccolò, and Carfagni, Monica. From 2D to 2.5 D ie from painting to tactile model. *Graphical Models*, 76(6):706–723, 2014.

[44] Giaccardi, Elisa. Collective storytelling and social creativity in the virtual museum: a case study. *Design issues*, 22(3):29–41, 2006. Publisher: MIT Press.

[45] Gibson, James J. Pictures, perspective, and perception. *Daedalus*, 89(1):216–227, 1960. Publisher: JSTOR.

[46] Gibson, James J. The ecological approach to the visual perception of pictures. *Leonardo*, 11(3):227–235, 1978.

[47] Gombrich, E. H. *Art and illusion : a study in the psychology of pictorial representation.* 2000.

[48] Gorman, Michael John. Art, optics and history: New light on the Hockney thesis. *Leonardo*, 36(4):295–301, 2003. Publisher: MIT Press.

[49] Grau, Oliver. *Virtual Art: from illusion to immersion.* MIT press, 2003.

[50] Guillou, Erwan, Meneveaux, Daniel, Maisel, Eric, and Bouatouch, Kadi. Using vanishing points for camera calibration and coarse 3D reconstruction from a single image. *The Visual Computer*, 16(7):396–410, 2000.

[51] Hagen, Margaret A. *Varieties of realism: Geometries of representational art.* CUP Archive, 1986.

[52] Hamill, Sam. *The Essential Chuang Tzu.* Shambhala Publications, 1999.

[53] Hao, Bing. Study on the Construction Techniques of Virtual Environment in Traditional Chinese Landscape Painting. Atlantis Press, 2018.

[54] He, Yeju 贺业钜. 中国古代城市规划史. 中国建筑工业出版社, 北京, 1996.

[55] Hearn, Maxwell K. *Cultivated Landscapes: Chinese Paintings from the Collection of Marie-Hélène and Guy Weill.* Metropolitan Museum of Art, 2002.

[56] Hearn, Maxwell K and Fong, Wen. *Along the Riverbank: Chinese Painting from the CC Wang Family Collection*. Metropolitan Museum of Art, 1999.

[57] Heelan, Patrick A. Toward a New Analysis of the Pictorial Space of Vincent Van Gogh. *The Art Bulletin*, 54(4):478–492, 1972. Publisher: Taylor & Francis.

[58] Hegel, G.W.F. *Aesthetics: Lectures on Fine Art*. Aesthetics: Lectures on Fine Art. Clarendon Press, 1975.

[59] Heidegger, Martin. Art and space. *Man and world*, 6(1):3–8, 1973. Publisher: Springer.

[60] Hochberg, Julian and McAlister, Edward. A quantitative approach, to figural" goodness". *Journal of Experimental Psychology*, 46(5):361, 1953. Publisher: American Psychological Association.

[61] Hockney, David and Falco, Charles M. Optical insights into Renaissance art. *Optics and Photonics News*, 11(7):52–59, 2000. Publisher: Optical Society of America.

[62] Hockney, David and Falco, Charles M. The Art of the science of Renaissance painting. 2004.

[63] Horry, Youichi, Anjyo, Ken-Ichi, and Arai, Kiyoshi. Tour into the picture: using a spidery mesh interface to make animation from a single image. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 225–232. ACM Press/Addison-Wesley Publishing Co., 1997.

[64] Hsi, Kuo. *An Essay on Landscape Painting (Lin Ch'üan Kao Chih)*. John Murray, 1935.

[65] Hu, Fuchen 胡孚琛. 道家和道教形、"气"、神三重结构的人体观. In 中国古代思想中的"气"论及身体观, pages 171–176. 台北: 巨流图书公司, 1993.

[66] Huhtamo, Erkki. *Illusions in motion: media archaeology of the moving panorama and related spectacles*. Mit Press, 2013.

[67] Hung, Wu. Reborn in paradise: a case study of Dunhuang sutra painting and its religious, ritual and artistic context. *Orientations*, 23:52–60, 1992.

[68] Hung, Wu. The Transparent Stone: Inverted Vision and Binary Imagery in Medieval Chinese Art. *Representations*, (46):58–86, 1994.

[69] Hurvitz, Leon. Tsung Ping's Comments on Landscape Painting. *Artibus Asiae*, 32(2/3):146–156, 1970. Publisher: JSTOR.

[70] Husserl, Edmund. *The idea of phenomenology: a translation of Die Idee Der Phänomenologie Husserliana II*, volume 8. Springer Science & Business Media, 2013.

[71] Kadar, Endre E and Effken, Judith A. Paintings as architectural space:"Guided Tours" by Cezanne and Hokusai. *Ecological Psychology*, 20(4):299–327, 2008.

[72] Kang, Hyung Woo, Pyo, Soon Hyoung, Anjyo, Ken-ichi, and Shin, Sung Yong. Tour into the picture using a vanishing line and its extension to panoramic images. In *Computer Graphics Forum*, volume 20, pages 132–141. Wiley Online Library, 2001.

[73] Kapoula, Zoï, Yang, Qing, Vernet, Marine, and Bucci, Maria-Pia. Eye movements and pictorial space perception: studies of paintings from Francis Bacon and Piero della Francesca. *Cognitive Semiotics*, 5(fall2009):103–121, 2009. Publisher: De Gruyter.

[74] Kemp, Martin. The science of art: Optical themes in western art from Brunelleschi to Seurat. 1990. Publisher: Yale University Press New Haven.

[75] Krikke, Jan. Axonometry: A matter of perspective. *IEEE Computer Graphics and Applications*, 20(4):7–11, 2000.

[76] Kuitert, Wybe. Composition of scenery in Japanese pre-modern gardens and the three distances of Guo Xi. *Studies in the History of Gardens & Designed Landscapes*, 33(1):1–15, 2013.

[77] Law, Sophia Suk-mun. Being in traditional Chinese landscape painting. *Journal of Intercultural Studies*, 32(4):369–382, 2011. Publisher: Taylor & Francis.

[78] Lescop, Laurent. 360 vision, from panoramas to VR. In *ENVISIONING ARCHITECTURE: SPACE/TIME/MEANING*, volume 1, pages 226–232. Published by the Mackintosh School of Architecture and the School of Simulation and Visualization at the Glasgow School of Art, 2017.

[79] Lescop, Laurent. Narrative grammar in 360. In *2017 IEEE International Symposium on Mixed and Augmented Reality Adjunct Proceedings*, pages 254–257, 2017.

[80] Lescop, Laurent and Lu, Yang. Les dimensions de la perspective, quand les modes de représentations européennes et chinoises rencontrent l'image numérique. In *SHS Web of Conferences*, volume 47, page 1015. EDP Sciences, 2018.

[81] Li, Jia, Nie, Jin-Wei, and Ye, Jing. Evaluation of virtual tour in an online museum: Exhibition of Architecture of the Forbidden City. *PloS one*, 17(1):e0261607, 2022. Publisher: Public Library of Science San Francisco, CA USA.

[82] Li, Nan and Huang, Zhiyong. Tour into the picture revisited. 2001.

[83] Li, Yanfeng 李彦锋. 卷轴与中国画叙事性刍议. 齐鲁艺苑*:* 山东艺术学院学报, (1):13–15, 2009.

[84] Liang, Sicheng 梁思成. 营造法式注释. 中国建筑工业出版社, 1983.

[85] Lin, Daoyu, Wang, Yang, Xu, Guangluan, Li, Jun, and Fu, Kun. Transform a simple sketch to a Chinese painting by a multiscale deep neural network. *Algorithms*, 11(1):4, 2018.

[86] Liu, Changlin 刘长林. 说"气". In 中国古代思想中的"气"论及身体观, pages 101–140. 台北: 巨流图书公司, 1993.

[87] Liu, WenYing 刘文英. 中国古代的时空观念. 南开大学出版社, 2000.

[88] Loran, Erle. *Cézanne's Composition: Analysis of his form with diagrams and photographs of his motifs.* Univ of California Press, 2006.

[89] Lotze, Hermann. *Microcosmus: An essay concerning man and his relation to the world*, volume 2. Books for Libraries Press, 1897.

[90] Lu, Yang and Lescop, Laurent. 3D Reconstruction Based on the Visual Structure of Chinese Hand-Scroll Painting. In *Vision 2021: Perception to Execution*, pages 174–182, Metro Atlanta, Georgia, USA, October 2021.

[91] Lu, Yang and Lescop, Laurent. Virtual Tour in the Chinese Hand-Scroll Painting. In *Monograph of the 15th Biennial International Conference of the European Architectural Envisioning Association*, pages 471–482, Huddersfield, United Kingdom, September 2021. University of Huddersfield.

[92] Ma, Wei, Kim, Hae Kwang, Wang, Yizhou, Gao, Wen, and Oh, Weon Geun. Binocular stereopsis of traditional Chinese paintings. In *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry*, pages 411–414. ACM, 2011.

[93] Ma, Wei, Liu, Yang, Wang, Yizhou, Xu, Yingqing, Zha, Hongbin, and Gao, Wen. Interactive viewpoint-space navigation for visual-audio exhibition of painting. In *Multimedia and Expo (ICME), 2010 IEEE International Conference on*, pages 540–545. IEEE, 2010.

[94] Ma, Wei, Wang, Yizhou, Xu, Ying-Qing, Li, Qiong, Ma, Xin, and Gao, Wen. Annotating traditional Chinese paintings for immersive virtual exhibition. *Journal on Computing and Cultural Heritage (JOCCH)*, 5(2):6, 2012.

[95] Maeda, Robert J. Chieh-Hua: Ruled-Line Painting in China. *Ars Orientalis*, pages 123–141, 1975. Publisher: JSTOR.

[96] Maruyama, Toshiaki. 中国古代"气"的特质. In 中国古代思想中的"气"论及身体观, pages 157–169. 巨流图书公司, 台北, 1993.

[97] McCall, Vikki and Gray, Clive. Museums and the 'new museology': theory, practice and organisational change. *Museum Management and Curatorship*, 29(1):19–35, 2014. Publisher: Taylor & Francis.

[98] McDowell, Maxson J. Pictorial Space throughout Art History: Cézanne and Hofmann. 2006. Publisher: Citeseer.

[99] Merleau-Ponty, Maurice. *The visible and the invisible: Followed by working notes.* Northwestern University Press, 1968.

[100] Merleau-Ponty, Maurice. *Phenomenology of Perception.* Routledge, 2012 edition, 2012.

[101] Needham, Joseph. *Science in traditional China: A comparative perspective.* Chinese University Press, 1981.

[102] Oettermann, Stephan and Schneider, Deborah Lucas. *The panorama: history of a mass medium*, volume 2. Zone Books New York, 1997.

[103] Osborne, Harold. *The Oxford companion to art.* Oxford University Press, USA, 1970.

[104] Packer, Randall and Jordan, Ken. *Multimedia: from Wagner to virtual reality.* WW Norton & Company, 2002.

[105] Pan, Zhigeng, Jiang, Ruiying, Liu, Gengdai, and Shen, Cailiang. Animating and interacting with ancient Chinese painting-Qingming festival by the riverside. In *Culture and Computing (Culture Computing), 2011 Second International Conference on*, pages 3–6. IEEE, 2011.

[106] Pan, Zhigeng, Liu, Gengdai, and Li, Zhi. Virtual presentation and animation of Qingming Festival by the riverside. In *Digital Media and its Application in Museum & Heritages, Second Workshop on*, pages 102–105. IEEE, 2007.

[107] Pang, Pu 庞朴. 一分为三论. 上海古籍出版社, 2003.

[108] Panofsky, Erwin, Wood, Christopher S, and Wood, Christopher. *Perspective as symbolic form*. Zone books New York, 1991.

[109] Peng, Feng 彭锋. 意境与气氛——关于艺术本体论的跨文化研究. 北京大学学报*(*哲学社会科学版*)*, 51(4):24–31, 2014.

[110] Pérez-Gómez, Alberto and Pelletier, Louise. Architectural representation beyond perspectivism. *Perspecta*, 27:21–39, 1992.

[111] Robinet, Isabelle. Visualization and ecstatic flight in Shangqing Taoism. *Taoist meditation and longevity techniques*, pages 157–90, 1989. Publisher: Center for Chinese Studies, University of Michigan Ann Arbor.

[112] Robinet, Isabelle, Kohn, Livia, and Yoshinobu, Sakade. Taoist Meditation and Longevity Techniques. 1989. Publisher: University of Michigan, Center for Chinese Studies.

[113] Saegusa, Tomomasa and Chikatsu, Hirofumi. 3D modeling and representation of "Ideal City" painted by Piero della Francesca. In *Proceedings of ISPRS Congress*, pages 12–23, 2004.

[114] Sasaki, Ken-ichi. Perspectives East and West. *Contemporary Aesthetics*, 11(1):16, 2013.

[115] Saxena, Ashutosh, Chung, Sung H, and Ng, Andrew Y. 3-d depth reconstruction from a single still image. *International journal of computer vision*, 76(1):53–69, 2008.

[116] Saxena, Ashutosh, Sun, Min, and Ng, Andrew Y. Make3d: Learning 3d scene structure from a single still image. *IEEE transactions on pattern analysis and machine intelligence*, 31(5):824–840, 2009.

[117] Schapiro, Meyer. *Theory and philosophy of art: Style, artist, and society*, volume 4. George Braziller New York, 1994.

[118] Schweibenz, Werner. The virtual museum: an overview of its origins, concepts, and terminology. *The Museum Review*, 4(1):1–29, 2019.

[119] Sdegno, Alberto. 3D Reconstruction of a Canaletto Painting. page 7.

[120] Seel, Martin. *Aesthetics of appearing*. Stanford University Press, 2005.

[121] Shi, Shengxun 时胜勋. 关于宗炳《画山水序》学术地位的十一个问题. 文艺研究, (3), 2011.

[122] Silbergeld, Jerome. *Chinese Painting Style: Media, Methods, and Principles of Form*. University of Washington Press Seattle and London, 1982.

[123] Silbergeld, Jerome. *Mind Landscapes: The Paintings of CC Wang*. Henry Art Gallery, 1987.

[124] Silbergeld, Jerome. Back to the Red Cliff: Reflections on the Narrative Mode in Earlyliterati Landscape Painting. *Ars Orientalis*, 25:19–38, 1995. Publisher: JSTOR.

[125] Sirén, Osvald. *The Chinese on the Art of Painting: Texts by the Painter-critics, from the Han Through the Ch'ing Dynasties*. Courier Corporation, 2013.

[126] Smith, Judith G and Fong, Wen. *Issues of Authenticity in Chinese Painting*. Metropolitan Museum of Art, 1999.

[127] Sullivan, Michael. *The birth of landscape painting in China*, volume 1. Univ of California Press, 1962.

[128] Sullivan, Michael. *Symbols of eternity: the art of landscape painting in China*. Stanford University Press, 1979.

[129] Sullivan, Michael. The Arts of China. *CA: University of California Press*, 2008.

[130] Sylaiou, Stella, Mania, Katerina, Paliokas, Ioannis, Pujol-Tost, Laia, Killintzis, Vassilis, and Liarokapis, Fotis. Exploring the educational impact of diverse technologies in online virtual museums. *International Journal of Arts and Technology*, 10(1):58–84, 2017. Publisher: Inderscience Publishers (IEL).

[131] Tang, Yijie 汤一介. 郭象与魏晋玄学. 北大名家名著文丛. 北京大学出版社, 2009.

[132] Taylor, B. and Jopling, J. *Dr. Brook Taylor's Principles of Linear Perspective, Or, The Art of Designing Upon a Plane the Representation of All Sorts of Objects: As They Appear to the Eye*. M. Taylor, 1835.

[133] Thomasson, Amie L. The Ontology of Art. In *The Blackwell Guide to Aesthetics*. 2004.

[134] Thompson, Seth. Cultural Heritage and Spectacle: Painted and Digital Panoramic Re-Presentations of Versailles. 2016.

[135] Tong, Qingbing 童庆炳. 文心雕龙物以情观说. 北京师范大学学报, 2011.

[136] Tseng, Lillian Lan-ying. Traditional Chinese Painting through the Modern European Eye—The Case of Ludwig Bachhofer. In *Tradition and Modernity, Comparetive Perspectives*, pages 508–513. Peking University Press, 2006.

[137] Tuan, Yi-Fu. *Space and place: The perspective of experience*. U of Minnesota Press, 1977.

[138] Tyler, Christopher W and Chen, Chien-Chung. Chinese perspective as a rational system: relationship to Panofsky's symbolic form. *Chin. J. Psychol*, 53:371–391, 2011.

[139] Vergo, Peter. *New museology*. Reaktion books, 1997.

[140] Wagemans, Johan, Van Doorn, Andrea J, and Koenderink, Jan J. Measuring 3D point configurations in pictorial space. *i-Perception*, 2(1):77–111, 2011. Publisher: SAGE Publications Sage UK: London, England.

[141] Wang, Bi 王弼. 周易略例. 王弼集校释, 609, 1980. Publisher: 北京: 中华书局.

[142] Wang, Bomin 王伯敏. 中国山水画的透视. 天津人民美术出版社, 1981.

[143] Wang, Eugene Y. Oneiric horizons and dissolving bodies: Buddhist cave shrine as mirror hall. *Art History*, 27(4):494–521, 2004.

[144] Wang, Linping, Zhang, Chi, Lu, Tingwen, and Zhang, Qing. Virtual Presenting a Famous Chinese Landscape Painting Based on Textured 3-D Model. In *2009 2nd International Congress on Image and Signal Processing*, pages 1–4. IEEE, 2009.

[145] Wang, Tingting, Mo, Lei, Vartanian, Oshin, Cant, Jonathan S, and Cupchik, Gerald. An investigation of the neural substrates of mind wandering induced by viewing traditional Chinese landscape paintings. *Frontiers in human neuroscience*, 8:1018, 2015. Publisher: Frontiers.

[146] Wang, Yangming 王阳明. 传习录. 岳麓书社, 长沙, 2004.

[147] Ward, John L. The perception of pictorial space in perspective pictures. *Leonardo*, 9(4):279–288, 1976. Publisher: The MIT Press.

[148] Wilhelm, H. and Baynes, C.F. *The I Ching or Book of Changes*. Bollingen Series. Princeton University Press, 2011.

[149] Wolfflin, Heinrich and Hottinger, Marie Donald. *Principles of art history*. 1929.

[150] Wu, Cong 吴葱. 在投影之外: 文化视野下的建筑图学究. 天津大学出版社, 2004.

[151] Wu, Hung. *The Double Screen: Medium and Representation in Chinese Painting*. Essays in Art and Culture. Reaktion Books, 1996.

[152] Wu, Hung. *"Space" in Art History*. Shanghai People's Press (上海人民出版社), 2018.

[153] Wu, Hung 巫鸿. 礼仪中的美术, volume 2. 生活读书新知三联书店, 2005.

[154] Wu, Hung 巫鸿. 图像的转译与美术的释读. 生活读书新知三联书店, 2006.

[155] Xiao, Jing. *Challenging cavalier perspective: an iconological study of visual perception of depth in Chinese representational space*. PhD Thesis, University of Nottingham, 2013.

[156] Xue, Yongnian 薛永年. 卷轴画史概说. 新美术, (3):43–46, 1993.

[157] Yang, Rubin 杨儒宾. 中国古代思想中的气论及身体观. 巨流图书公司, 台北, 1993.

[158] Ye, Lang 叶朗. 中国美学史大纲. 上海人民出版社, 1985.

[159] Ye, Lang 叶朗. 美学原理. 博雅大学堂. 北京大学出版社, 2009.

[160] Ye, Lang 叶朗and Zhu, Liangzhi 朱良志. 中国美学通史. 江苏人民出版社, 2014.

[161] Yu, JinHui, GuoMing, Luo, and QunSheng, Peng. Image-based synthesis of Chinese landscape painting. *Journal of Computer Science and Technology*, 18(1):22–28, 2003.

[162] Zhang, Songhai, Chen, Tao, Zhang, Yifei, Hu, Shimin, and Martin, Ralph. Video-based running water animation in Chinese painting style. *Science in China, Series F: Information Sciences*, 52(2):162–171, 2009.

[163] Zhou, Pengbo, Li, Kaiyue, Wei, Wei, Wang, Zhe, and Zhou, Mingquan. Fast generation method of 3D scene in Chinese landscape painting. *Multimedia Tools and Applications*, pages 1–17, 2019.

[164] Zhou, Zhenfu 周振甫. 周易译注. 中华书局, 2018.

[165] Zhu, Guangqian 朱光潜. 西方美学史. 中华现代学术名著丛书. 商务印书馆, 2011.

[166] Zhu, Liangzhi 朱良志. 中国美学十五讲. 名家通识讲座书系. 北京大学出版社, 2006.

[167] Zong, Baihua 宗白华. 中西画法所表现的空间意识. 安徽教育出版社, 合肥, 1994.

[168] Zong, Baihua 宗白华. 美学散步, volume 1. 上海人民出版社, 2000.

[169] 巫鸿. 时空中的美术. 生活. 读书. 新知三联书店, 2016.

Appendix A

Chronology of Dynastic China

Xia Dynasty 夏	c.2100 - c.1600 BC
Shang Dynasty 商	c.1600 - c.1100 BC
Zhou Dynasty 周	c.1100 - 256 BC
Western Zhou 西周	c.1100 - 771 BC
Eastern Zhou 东周	770 - 256 BC
Spring and Autumn Period 春秋	770 - 481 BC
Warring States Period 战国	481 - 221 BC
Qin Dynasty 秦	221 - 206 BC
Han Dynasty 汉	206 BC - AD 220
Western Han Dynasty 西汉	206 BC - AD 9
Wang Mang Interregnum 王莽新朝	9 - 23
Eastern Han 东汉	25 - 220
Six Dynasties 六朝	220 - 589
Three Kingdoms 三国	220 - 265
Western Jin Dynasty 西晋	265 - 317
Southern Dynasty 南朝	317 - 589
Northern Dynasty 北朝	386 - 581
Sui Dynasty 隋	581 - 618
Tang Dynasty 唐	618 - 907
Five Dynasties 五代	907 - 960
Liao Dynasty 辽	916 - 1125
Song Dynasty 宋	960 - 1279
Northern Song 北宋	960 - 1127
Southern Song 南宋	1127 - 1279
Jin Dynasty 金	1115 - 1234
Yuan Dynasty 元	1272 - 1368
Ming Dynamic 明	1368 - 1644
Qing Dynasty 清	1644 - 1991
Republic 中华民国	1912 - 1949
People's Republic 中华人民共和国	1949 -

Appendix B

Projection transformation

B.1 Transformation of oblique projection

Assume that, $Q = (x, y, z)$ is a point in space whose distance to the projection plane is $z_v p$, $P = (x_p, y_p, z_p)$ is the oblique projection of P, $R = (x, y, z_v p)$ is the front projection of Q, θ is the angle between PR and the X-axis. [Figure B.1.1]

Then,

$$x_p = x + PR\cos\theta$$
$$y_p = y + PR\sin\theta \tag{B.1}$$

The length of PR depends on the angle α between the projection ray and the projection plane, as well as the distance $z_v p$.

$$PR = \frac{z_v p - z}{\tan\alpha} \tag{B.2}$$

B.2 Transformation of perspective projection

Assume that, $Q = (x, y, z)$ is a point in space, $P = (x_p, y_p, z_p)$ is the perspective projection of P, n is the distance from the viewpoint to the near clipping plane. [Figure B.2.1]

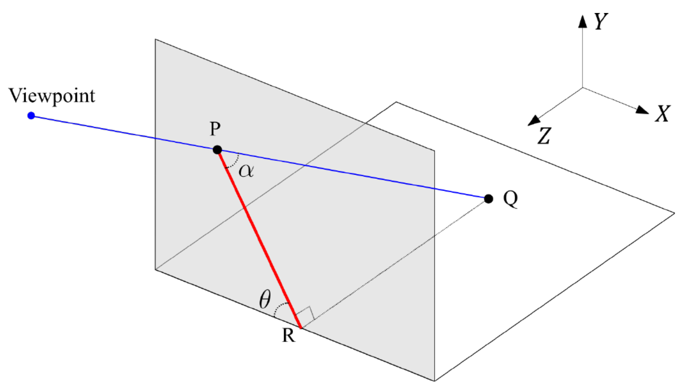

Figure B.1.1: Transformation of oblique projection

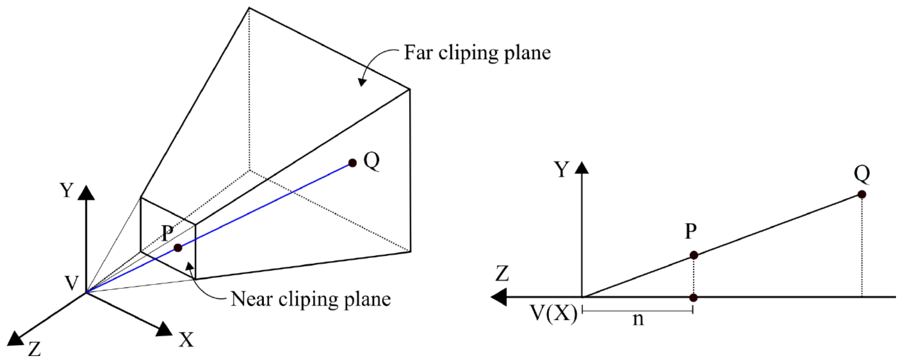

Figure B.2.1: Transformation of perspective projection

According to the similar triangles,

$$\begin{aligned} \frac{x_p}{n} &= \frac{x}{z}, \\ \frac{y_p}{n} &= \frac{y}{z}. \end{aligned} \tag{B.3}$$

Then,

$$\begin{aligned} x_p &= \frac{nx}{z}, \\ y_p &= \frac{ny}{z}, \\ z_p &= n. \end{aligned} \tag{B.4}$$

List of Figures

Titre : VISITE DANS LA PEINTURE: Conception d'un système pour l'exposition virtuelle de peintures chinoises sur rouleau

Mots clés : Exposition virtuelle, Réalité virtuelle, Peinture chinoise à rouleau, Structure visuelle

Résumé : La peinture chinoise à rouleau (CHSP) est une forme de peinture chinoise typique. La manière correcte de regarder une CHSP est de la faire défiler à la main. Regarder une CHSP est considéré comme une expérience fantastique de visite dans son monde diégétique.

Dans les expositions actuelles de CHSP, ces points ne sont pas bien représentées. Premièrement, la plupart des CHSP sont présentés dans des boîtes en verre. Les spectateurs ne peuvent pas les manipuler en les faisant défiler. Deuxièmement, l'expérience de la visite d'une peinture n'est interprétée que par de simples annotations. Les spectateurs ne peuvent pas acquérir une expérience complète à partir de ces annotations.

Cette recherche a exploré l'utilisation de la réalité virtuelle (RV) pour combler ces lacunes. L'objectif de cette recherche est de développer un système d'exposition basé sur la RV, qui peut simuler de manière synchrone l'expérience de visualisation de l'ancien spectateur dans le monde réel et l'expérience de visite dans le monde diégétique.

Le système est présenté comme une application de RV qui comprend une CHSP interactive qui peut être manipulé selon les principes originaux. De manière synchrone, l'utilisateur de la RV sera déplacé en fonction de son point de focalisation sur la CHSP virtuelle, puis un monde diégétique englobant et aléatoire sera construit.

Title : TOUR INTO PAINTING: System Design for Virtual Exhibition of Chinese Hand-Scroll Painting

Keywords : Virtual exhibition, Virtual Reality, Chinese Hand-Scroll Painting, Visual structure

Abstract : The Chinese Hand-Scroll Painting (CHSP) is a typical Chinese painting form. The proper way to view a CHSP is scrolling it by hands. Watching a CHSP is considered as a fantasy experience of touring in its diegetic world.

In current exhibitions of CHSPs, these points are not well represented. First, most of CHSPs are presented in glass boxes. Viewers can not manipulate it in scrolling manner. Second, the experience of touring into painting is only interpreted by simple annotations. Viewers can not gain a full experience based on them.

This research aims to implementing these shortcoings by developing a VR-based exhibition system, which can synchronously simulate the ancient viewer's viewing experience in the real world and the touring experience in the diegetic world.

The system is presented as a VR application that includes an interactive CHSP that can be manipulated according to the original principles. In a synchronous way, the VR user will be moved according to his focus point on the virtual CHSP, and then, an encompassing, random diegetic world will be built.