

The Construction of Sichuan Image under Multimodal Visual Grammar—Taking the Documentary “Aerial China (Sichuan)” as an Example

Xinxin Hong, Cheng Duan

College of Foreign Languages and Cultures, Chengdu University of Technology, Chengdu, China
Email: 798450661@qq.com

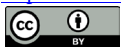
How to cite this paper: Hong, X. X., & Duan, C. (2020). The Construction of Sichuan Image under Multimodal Visual Grammar—Taking the Documentary “Aerial China (Sichuan)” as an Example. *Open Journal of Social Sciences*, 8, 108-120.
<https://doi.org/10.4236/jss.2020.812010>

Received: November 17, 2020

Accepted: December 12, 2020

Published: December 15, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Based on the visual grammar raised by Gunther Kress and Theo van Leeuwen, this paper tasks visual text analysis as the research method, and explores the dynamic image construction of Sichuan Province from three dimensions of “Sichuan characteristics”, “representative meaning”, “interactive meaning” and “compositional meaning” by analyzing the image modals in the national geographic documentary “Aerial China (Sichuan)”.

Keywords

Visual Grammar, Multimodal, Documentary, National Image

1. Introduction

What is realized in today’s world is multi-modal presentation, in which the masses integrate symbolic resources to realize communication and exchange. Multi-modal discourse integrates a variety of social symbols to deepen the visual, auditory, tactile and other sensory experiences. As a typical multi-modal discourse, documentary is composed of language, image, sound and other modalities. It effectively combines text and image modalities to bring rich sensory experience to the audience and form a strong visual memory in them.

The documentary “Aerial China” produced by CCTV uses a full aerial image narrative to show China’s beautiful natural landscapes and colorful ecological environment, highlighting the brilliant achievements of China’s economic construction and presenting China’s image from a unique bird’s-eye perspective. It’s

a kind of popular and warm visual text material. Based on the visual grammar of Kress and van Leeuwen, this paper uses visual text analysis as a research method to make a qualitative analysis of the multimodality in the documentary, interpret the meaning of frames and analyze the hidden image construction.

2. Research Status of Multimodal Discourse Analysis

Since the 1990s, the multimodal turn of functional linguistics and discourse analysis has made multimodal discourse analysis one of the hot topics in linguistics and communication research. The theoretical basis is Halliday's functional theory of linguistic system. Social symbolists have established semantic grammar systems to describe symbols such as visual images (Kress & van Leeuwen, 1996). In "Reading Images: The Grammar of Visual Design (Narrative Representation)", Kress and van Leeuwen (1996) focus on the three main functions of language in system functional linguistics and divide the multimodal meaning system into representation, interaction and construction, which corresponds to the concept, interpersonal and group meaning of the language, summing up the framework of visual grammar analysis and developing social semiotics based on visual grammar. In "Visual Analysis Manual", Leewen & Jewitt (2001), introduce visual analysis in anthropology, cultural studies, psychoanalysis, ethnological studies, and film and television studies, as well as image content analysis and socio semiotics analysis.

International studies show new trends, reflected in the enhanced interdisciplinary multimodal study, increasingly mature critical discourse analysis, and the interpretation of non-linguistic symbols based on cognitive theory, library analysis or psychological empirical research.

Multimodal research has also made some progress in China. Since Li Zhanzi (2003) in "Social Semiotic Analysis of Multimodal Discourse" first introduced Kress and van Leeuwen's multimodal visual grammar theory to China, Chinese scholars have carried out extensive and in-depth discussion of on advertising discourse, news reports, television movies, classroom teaching, natural conversation and other multimodal forms from the system functional visual grammar, multimodal metaphor, multimodal corpora, multi-reading and writing and other perspectives.

As far as research methods are concerned, researchers have begun to use digital technology to annotate and simulate complex multimodal texts establish multimodal corpus and develop multimodal retrieval software (Baldry & Thibault, 2008; Gu 2006). Meanwhile, empirical research on viewers' cognition of multimodal discourse has been gradually rising through questionnaire survey, eye movement experiment and even brain imaging technology (Gidlof et al., 2012; Muller et al., 2012).

In terms of application, relevant researches focus on static media such as print ads, political comics, posters, foreign language teaching materials, photographic images, and dynamic media such as TV advertisements, films and propaganda

films. Although there are considerable studies on multimodal discourse analysis, there is a little research on the meaning structure of the national geographic image documentaries. In this paper, we will take the single-frame image as the analysis unit to integrate the image and discourse meaning resources in the anthology of “Aerial China (Sichuan)” under the framework of visual grammar from the shooting techniques and shooting angles, exploring the significance of multimodal documentary discourse presentation.

3. Multimodal Discourse Analysis of Documentaries

According to Halliday’s (1978) social semiotics theory, both linguistic and non-linguistic symbols cannot be regarded as invariable semantic codes, but as resources constructing meanings in certain contexts. Therefore, the visual “grammar” of Kress & Van Leeuwen (1996) is only a systematic description of image ideographic resources, rather than a rigid rule.

3.1. Analysis of the Representational Meaning

Kress (2010) defines modality as “the symbol resources that create meanings in social culture” and believes that “any modality (such as image, gesture, music) is a complete ideographic system which contains expression plane, lexicogrammar and discourse semantics like language”.

Representational meaning is the foundation of multimodal discourse construction, corresponding to the conceptual function of Halliday’s language metafunction. The representational meaning of image is composed of two parts: narrative representation and conceptual representation. Symbol resources can objectively feed back the modalities of the real world and the relationships between things. In the visual grammar, the representation modal presents the unfolding actions, the process of events and the transformative spatial arrangement.

The narrative representation includes action, reaction, speech and mental process.

In the course of action, elements form diagonal lines, and usually strong diagonal lines form vectors. The narrative vector undertakes the interaction of single or multiple participants, and is also an intermediary between the constituent elements. In this process, the actor sends out a vector signal and participates in it, which highlights the actor’s status. When participants are connected by the vector, it means that they do something for others or each other.

The vector component element of the reaction process is the gaze of the participants. The gaze vector has certain directionality. The gaze of the active participant points to that of another participant. The former is referred to as “reactor” and the latter “phenomenon”.

The documentary “Aerial China (Sichuan)” shows humanistic Sichuan through narrative representation. The face-changing performance is a major feature of Chengdu’s scenic spots, and also a major business card of Sichuan culture.

The narrative meaning of **Figure 1** is that “face-changing performance is loved by the audience”, echoing the language narrative that “Sichuan opera is the ‘Sichuan cuisine’ in Chinese opera. Face-changing is a unique skill in Sichuan opera. The faces of joy, anger, sadness and happiness can be switched instantly”. The close-up shooting of face-changing performances and fire-spraying and the representation of the audience’s reaction constitute the intransitive action process of the performer and the reaction process of the audience. The audience sitting under the stage and watching the face-changing performance on the stage constitutes the vector of the reaction process in the narrative representation (“audience watching the performance”), and the “watching” process is highlighted through their gaze. By taking close-up shots of the process of face-changing, the face-changing performer is placed in the center of the picture and becomes the most prominent participant, which arouses the psychological prominence of the documentary viewers. The narrative meaning here is to highlight that the face-changing of Sichuan opera is a historical and cultural image of Chengdu in Sichuan.

Leaving Chengdu, aerial photography takes the audience to Beichuan, a new homeland rebuilt after the earthquake. This set of image modalities takes the post-disaster reconstruction as the topic and recalls people’s special memory of the earthquake occurring in 2008 via vision. First, the clock of Xuankou Middle School damaged in the earthquake is brought into view from an overlooking perspective. This image is the conceptual reappearance in the meaning of representation. It is a symbolic process. There is no vector in the picture, but a commentary of the symbolic ruins, “the shocking scene of the big earthquake is still fixed at the site of Xuankou Middle School”. The deep meaning is that the 2008 Wenchuan Earthquake severely damaged Beichuan and ruined the former homeland.

Figure 2 and **Figure 3** present the life scenes of middle school students in New Beichuan Middle School in Beichuan’s newly built county after the earthquake, recreating narratives for the audience, and interpreting the meaning that



Figure 1. Face-changing of Sichuan Opera. (Aerial China (Season 2)—Sichuan)



Figure 2. Post-disaster reconstruction. (Students work hard)



Figure 3. Post-disaster reconstruction. (Students work hard)

“after brand-new planning and construction, the current Yingxiu has been re-born”. These two figures respectively show students watching the national flag rising and studying in the classroom. The gaze of students on the national flag, the watching and listening to the teacher of them in the classroom make the gaze constitute the vector of the reaction process. The meaning of the image is that “the figure of students studying hard for progress has become a symbol of the rebirth of New Beichuan”. The representation of the lives of local residents in their new homes rebuilt after the disaster is a symbolic process, symbolizing “the great power of Chinese nation’s spirit for earthquake relief” and that “disaster makes the Chinese nation more cohesive”. As an ideographic resource, this set of images continuously describes the indomitable will of the nation and the subjective initiative of the people in the face of merciless natural disaster.

Sichuan is a large province involving many ethnic groups and the largest settlement area of Yi people. Daliang Mountain is the main settlement area of Yi people in China. The scene in **Figure 4** is the grandest representation of the form in the film, presenting the audience with a bonfire carnival and a grand Yi



Figure 4. The Torch Festival of Yi people. (Will select the kindest and the wisest)

beauty pageant. The first picture shows that, in the wide field, hundreds of Yi women are arranged in two circles under the leadership of the singer, turning the yellow butter cloth umbrellas and walking slowly in the center of the square. The yellow butter cloth umbrella forms a color contrast with the green field with high color saturation, and the far lens presents a panoramic view. The high color saturation and central composition make the Yi girls arranged in a circle the most prominent participants. This set of images includes both the action process in the narrative representation of the Yi women's singing and dancing to celebrate the Torch Festival, and the symbolic suggestion process of the conceptual representation of the status of the Yi people's Torch Festival. The carnival of the Bonfire Festival and the highlight, the beauty pageant of the Yi people are used to symbolize that "this (Torch Festival) is the grandest traditional festival of the Yi people". The overall narrative significance of these frames is the artistic sentiment and traditional national cultural charm of the ethnic Yi people in production and life, showing the living status in the ethnic minority areas represented by the Yi people in Sichuan which is a major ethnic province.

Sichuan is also home to the Tibetan ethnic minority. Founded in 1792 in Ganzi Prefecture, Sichuan, Dege Sutra Printing House is the largest Tibetan sutra printing house in China. **Figure 5** shows the oldest engraving printing process by dark-skinned Tibetan wearing felt hats and Tibetan costume with the dry pages. This group of frames has narrative representation—"Tibetans are printing and drying the printed Tibetan pages". On the whole, it is conceptual representation, corresponding to the process of symbolic suggestion in language grammar. The engraved printed scriptures symbolize the traditional Tibetan intangible cultural skills and the time-honored precious cultural traditions, and the simple Tibetans symbolize the pure and meticulous craftsmanship of the Tibetans.

3.2. Interpretation of the Interactive Meaning

Interactive representation corresponds to the interpersonal function of Halliday's



Figure 5. Dege sutra printing house in Ganzi, Sichuan.

language metafunction, which is divided into four dimensions: contact, social distance, attitude and modality. Interactive meaning refers to the relationship of the image maker, the viewer and the various models in the image. The above three are also important elements of image meaning. Interactive meaning is a two-way interactive model, with one end of participants of the image, and the other end of the viewer. Information is transmitted at both ends, and the viewer can get the most optimized interaction.

The reference standard of attitude dimension is “view angle”, which expresses the meaning of “intervention” and “power” through five “view angles” in frontal and vertical directions. In the horizontal direction, it is divided into frontal angle and oblique angle to determine the degree of “intervention” in the image. In the vertical direction, it is divided into three dimensions: “low angle”, “level angle” and “high angle” to determine the attitude of the image viewer towards the image participants.

In the documentary visual text, the lens is dynamic, and the horizontal angle of view is determined by the shooting technique. The horizontal view angle means that the front lens depicts the image to get the viewer involved in the image scene; the horizontally oblique view angle creates distance between the viewer and the image participants. As for the vertical view angle, the shooting techniques of the documentary include overhead shooting and overhead shooting, top shooting, fixed field shooting, long-range shooting and aerial shooting.

Giant pandas are national treasures and “national business cards”, while Sichuan is the famous hometown of pandas. Through the choice of view angles, the film realizes the interaction between the viewer and the panda, and strengthens the impression on people. When the lens presents the scenes of panda, as shown in **Figure 6**, it uses the frontal shooting angle. According to the visual grammar, “frontal angle” means viewing participants in the image from an intervening perspective. “In the Panda Kindergarten, the little pandas are chasing their male nurses around and playing the classic game of holding thighs again.”

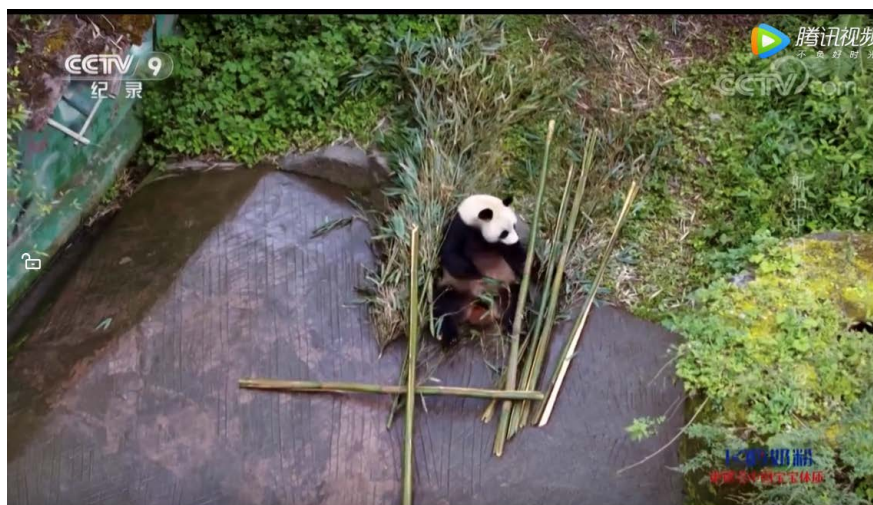


Figure 6. National treasure—panda.

Pandas run after their nurses, which is the action process of the narrative representation meaning. The frontal angle makes the viewer involved in the same scene of playing with the panda, forming a connection between, the viewer and the panda and narrowing the distance between the viewer and the cute panda in the film, thus making the viewer immersed in the image. The representation of panda life in the Wolong Conservation Park shows the significance of the national treasure panda being taken good care of by human beings in the protected area, and shows Sichuan's efforts in giant panda protection and research and ecological civilization construction.

Sichuan also has many famous tourist attractions. As the tallest stone Buddha stature in the world, the Leshan Giant Buddha connotes Chinese Buddhist culture. In the documentary, far shooting, low-angle shooting and close-up techniques are used to capture the Leshan Giant Buddha. In **Figure 7**, the Giant Buddha the largest Maitreya Buddha in the world, with its head in line with the top of the mountain, feet on the river, hands on the knees as well as a well-proportioned body and solemn expression—"It is 71 meters high, with 6-meter-long ears, 3.2-meter-long nose, 2.46-meter-wide left eye and 2.45-meter-wide right eye." The close-up of the giant Buddha's head enables the audience to have a close look at the Buddha. The combination of far and close lens enhances the audience's sense of interaction and realizes the interaction between images and viewers. The low-angle lens from the bottom to the up presents the Giant Buddha to the viewer. At this time, the image of the Buddha is awe-inspiring, while the viewer is in a weak position, highlighting the Buddha in the composition. The Leshan Giant Buddha is a rare stone sculpture, and because of the cultural cognition of Buddhism and the natural awe of the Buddha statue in the traditional Chinese thought, the effect of the low-angle shooting resonates with the inner cognition of the viewer.

The magnificent but precipitous Sichuan scenery is a result of hundreds of millions of years of geological movement. As a natural landscape, the precipitous

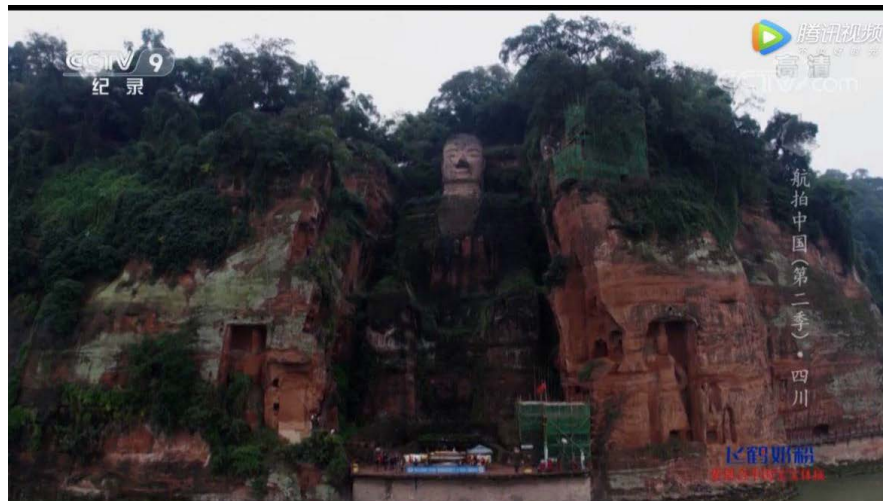


Figure 7. Leshan Giant Buddha.

landform is the geographical condition for breeding Sichuan culture. The interactive significance of image modal is achieved through the choice of the shooting techniques. Great mountains, gorges and rivers are shown many times in this film lasting for 50 minutes, including Songpan plateau, Hengduan Mountains, Qincheng Mountain, Emei Mountains, ridge, Siguniang Mountain (Queen Peak and Camel Peak), Gongga Mountains, Xiannaier Snow Mountain, Yangmai Yong Mountain and Xianuodduoji Snow Mountain. The film uses establishing shots to capture the snow mountain in Daocheng Yading of Sichuan (**Figure 8(a)**), creates wild views and shows the geographic environment and the scales of these mountains, making the viewer feel the extraordinary momentum of Western Sichuan snow mountains. At the same time, it expands multiple view angles and takes an aerial perspective to compensate for the limitations of people's perspective, letting the audience understand the actual spatial distance between the camera and the towering mountains. The three snow-capped mountains are closely linked behind the Hengduan Mountains, while the Xiannaier Peak towering into the clouds (**Figure 8(b)**), the Yangmai Yong Peak is shaped like a sword, and the Xiasongduoji Peak is as sharp as sawteeth. The camera moves from the panoramic view to the distant view, showing the relationship of the three mountains and the spatial jump gives the viewer a sense of novelty and creates a grand scene. When the camera is focused on the peaks, the long shot is used, which makes the viewer feel as if they are on the snow-capped mountains and peaks, reinforcing their impression of many towering mountains of Sichuan coexisting with the paradise on the earth.

3.3. Interpretation of the Compositional Meaning

Kress and Leeuwen also proposed a model of meaning integration and combination between image representation elements and interactive elements—"the composition meaning", which corresponds to the textual function of Halliday's metafunction. The viewer of the image focuses on the spatial layout of the dynamic



(a)



(b)

Figure 8. Siguniang Mountain.

model to establish the overall tone, which is helpful to judge the overall composition. There are three parts in composition representation: information value, salience and framing. “Information value” means that image elements realize their roles through their composition and placement position in the whole picture. Different information values are implied in the arrangement of image composition, which has a certain impact on human visual cognition.

Among them, salience is to attract the attention of viewers through a series of methods. In the documentary visual text, the object is emphasized through the relative size of the image participants, sharpness of the picture focus, color contrast, visual position, perspective angle and special cultural symbols.

In the documentary, the color contrast and the top shot are combined to connect the representation meaning and the interaction meaning to present the composition meaning. **Figure 9** shows the chili red of Pixian thick broad-bean sauce. The chili sauce in the jar shows different shades of red. The top shot shows the sorted jar, and the deep and light red is interlaced in the picture. It attracts the attention of the audience with bright red, strengthens the visual effect, and enhances people’s memory of Pixian thick broad-bean chili.

Towering mountains and deep valleys are the world’s unique landscapes. The

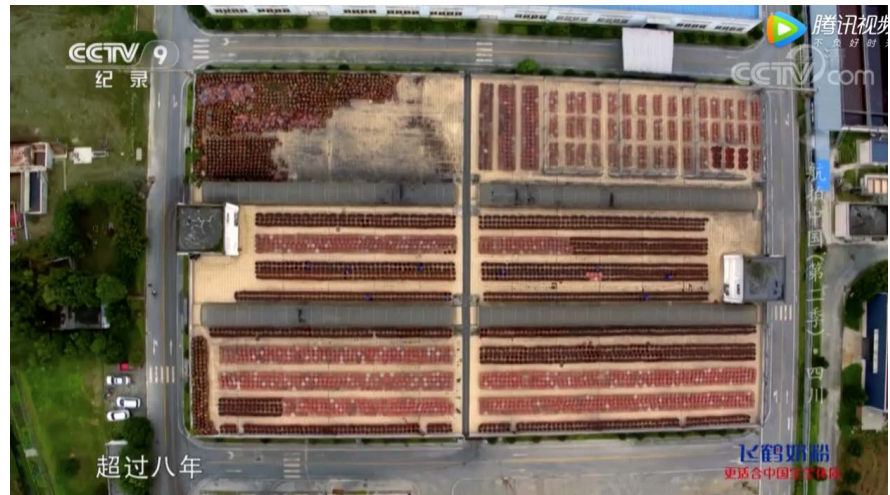


Figure 9. Pixian thick broad-bean sauce.

documentary uses the technique of color contrast to highlight the color characteristics of the images, which is the highlight application in the compositional meaning. Huanglong Scenic and Historic Interest Area is a world-class calcified landscape. Viewed from the sky, the natural calcified beach is majestic with golden water flow, “looking like a long golden dragon” (Figure 10). The lens moves from left to right, captures the calcified beach and finally presents the panorama of it, which is in line with people’s visual perception.

In Figure 11, the lens of Scenic Spot of Daocheng Yading uses color contrast. The golden meadow and colorful forest are interlaced, which attracts the attention of the audience. The documentary also uses spot photography to finish composition. In order to show the seasonal changes of Jiuzhaigou’s natural scenery, the documentary uses spot photography to show the change from colorful spring to snow-covered winter. Through the change of composition and visual position, the documentary presents images in various forms and shows various Sichuan wonders via the composition meaning, which constitutes the narration of natural wonders of Sichuan.

The analysis of this paper regards the documentary image modalities and commentary text modalities as language-like ideographic resources, interprets the meaningful representative images from the representational meaning, conceptual meaning and interactive meaning under the framework of visual grammar in the construction of the image of Sichuan, and analyzes the meaning and connotation of the overall image presentation, thus exploring the image of Sichuan in the multi-modal interaction of the documentary “Aerial China (Sichuan)”.

4. Conclusion

In this paper, the documentary is a complete dynamic multimodality discourse. The image modal and commentary text modal are regarded as resources with ideographic function. Visual grammar is used to analyze the integration of

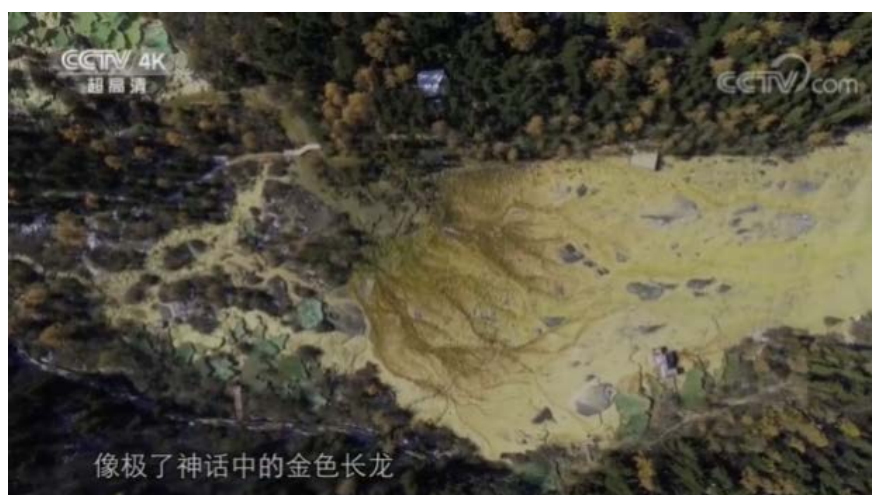


Figure 10. Huanglong, Sichuan.



Figure 11. Daocheng Yading Scenic Spot.

“narrative meaning”, “representation meaning” and “interactive meaning” in multiple scenes of the documentary, which interprets the overall narrative significance and image of “Aerial China (Sichuan)”. This paper is intended to enrich the perspective of multimodal discourse analysis and enhance the interdisciplinary nature of multimodal research, that is, to analyze documentaries with linguistic and semiotic knowledge and film and television media knowledge, and to investigate the significance of continuous visual narrative discourse and provide the multimodal research with certain inspiration by taking the documentary as a whole and complete multimodal discourse, combining pictures with commentary and breaking through the limitation of single image.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- Baldry, A., & Thibault, P. (2008). Application of Multimodal Concordances. *Journal of Language and Communication Studies, 41*, 12-41.
- Gidlof, K., Holmberg, N., & Sandberg, H. (2012). The Use of Eye Interviews to Study Teenagers' Exposure to Online Advertising. *Visual Communication, 11*, 329-345. <https://doi.org/10.1177/1470357212446412>
- Gu, Y. G. (2006). Multimodal Text Analysis: A Corpus Based Approach to Situated Discourse. *Text and Talk, 26*, 127-167. <https://doi.org/10.1515/TEXT.2006.007>
- Halliday, M. A. K. (1978). *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Arnold.
- Kress, G. (2010). *Multimodality*. London: Routledge.
- Kress, G., & Van Leeuwen, T. (1996). *Reading Images: The Grammar of Visual Design (Narrative Representation)* (pp. 1-29). London: Routledge.
- Leeuwen, T. V., & Jewitt, C. (Eds.) (2001). *Handbook of Visual Analysis*. London: Sage.
- Li, Z. Z. (2003). Social Semiotics Analysis of Multimodal Discourse. *Foreign Language Studies, 5*, 1-8.
- Muller, M. G., Kappas, A., & Old, B. (2012). Perceiving Press Photography: A New Integrative Model, Combining Iconology with Psycho Physiological and Eye-Tracking Methods. *Visual Communication, 11*, 307-328. <https://doi.org/10.1177/1470357212446410>