Video–Geographic Scene Fusion Expression Based on Eye Movement Data

Xiaozhi Wang
College of Information Engineering, Jiangsu Key Laboratory of Grain Big-data Mining
Nanjing University of Finance & Economics
Nanjing, China
1120190818@stu.nufe.edu.cn

Yujia Xie*
College of Information Engineering, Jiangsu Key Laboratory of Grain Big-data Mining
Nanjing University of Finance & Economics
Nanjing, China
9120181003@nufe.edu.cn

Xing Wang
Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education
Nanjing Normal University
Nanjing, China
181301021@njnu.edu.cn

***Abstract*—Video–geographic scene fusion is the development trend in surveillance video systems. How to analyze the influence of various factors on the fusion expression effect is of great significance. In this paper, we select video compression state, video scene fusion expression and viewer features as the impact factors of fusion expression visualization. And the influence of above factors was analyzed systematically by eye movement experiment. Experimental results show that these three factors have a certain impact on the video–geographic scene fusion expression effect. The expression effect of uncompressed video is better than that of compressed video. No significant difference exists between the embedded and foreground projection, but a significant difference exists between the correlation and the previous two expressions. In terms of viewer features, we take gender factors into account in this paper. Under the condition of normal corrected vision and the same age, the accuracy of information received by women is higher than that of men based on the gender analysis of the observer.**

Keywords—video–geographic scene fusion expression, parametric model, eye movement experiment, eye movement data analysis, visualization effect analysis

I. Introduction

With the continuous development of urban video monitoring technology, surveillance cameras are installed increasingly in public places and areas in cities, such as streets, shopping malls, and other places. Surveillance cameras can record what happens in their field of vision at any time and it can provide a real-time basis for urban security and construction. However, the shooting range of a single camera is limited and thus insufficient to cover an entire monitoring area. Therefore, multiple cameras are needed to cover the entire area. In a traditional monitoring system, the security personnel of the control center monitor the security area through a screen wall (as shown in Fig. 1). The spatial relationship between camera views is not intuitive in adjacent views, which increases the visual pressure on the monitoring personnel. It has received significant research interest and made great progress of the fusion expression between video monitoring information and the geographical scene [1]–[4].

Video-scene fusion expression refers to the integrated expression of video images and geospatial information. Specifically, it is placing the video information obtained by different cameras in a unified geographic scene model. Video-scene fusion expression can help monitoring personnel understand video contents effectively combined with geospatial information. However, under different space–time constraints and information expression requirements, the visualization effect of video-scene fusion expression has not been analyzed systematically and effectively. And the advantages and disadvantages of different information expression modes and strategies have not been compared comprehensively quantitatively. To solve these problems, this study attempts to analyze video-scene fusion information expression based on eye movement data. To study the effect of information fusion with surveillance videos in the corresponding scene, the video network composed of geographical cameras is used. The study is also combined with the layout of 2D and 3D scenes and the theoretical and practical cognition of the development process of eye movement technology. This paper proposes an analysis method related to eye movement data, and achieves the goal of evaluating the effect of video–geographical scene information fusion expression.

Fig. 1. Multi-screen traditional monitoring system.

II. Related Work

A. Fusion Expression of Video and Geographic Scene

Geographic video is a new way of integrating geographic information and video clips for geospatial expression. Scholars at home and abroad have explored the theory and application of video and geographic information system (GIS) integration. In 1978, Professor Lippman of MIT integrated video and spatial data for the first time and developed dynamic and interactive hypermedia maps [5]. Then, multimedia technology

978-1-6654-2309-0/21/$31.00 ©2021 IEEE

has been gradually introduced into the GIS field, and multimedia map, multimedia GIS, and geographic hypermedia concepts have been proposed. In recent years, some foreign scholars have carried out related research by combining video and geographic information: proposed the framework of video map and designed the conceptual scheme of data acquisition, processing, and application in the field [6]; established the geographic index of video clips and realized the related data acquisition and application system [7]–[10]; carried out spatiotemporal modeling of object entities in video images and constructed video metadata to complete the mapping between the spatial position and the video frame [11], [12]. Meanwhile, Chinese scholars have also carried out relevant research, such as the application of video and geographic information integration in railway and highway visualization management [13]–[15], the design of the vehicle mobile video monitoring system based on maps [16], the research and development of a portable positioning video system [17], the design of a geographic video data model and the application of geographic video in the network environment [18]–[20].

B. *Analysis of Eye Movement Data*

Eye tracking was first applied in psychology and neurobiology and later applied in many geography fields, such as cartography and GIS [21]. With the development of eye movement experimental equipment and the diversification of eye movement data analysis methods, eye movement experiments and data analysis have been applied to 2D maps and 3D scenes and in watching videos to analyze the psychological activities of users in various environments.

Eye movement experiment is used in 2D maps. In 2013, Weihua D et al. [22] proposed the application of eye movement data in evaluating the performance of dynamic maps. In 2016, Brychtova A et al. [23] used remote eye tracking equipment to obtain user eye movement data, analyzed the influence of color distance and font size on map readability, and evaluated the success rate, efficiency, and flicker frequency. For 3D geographic scenes, Stanislav et al. [24] studied users' perception of 2D and 3D terrains through eye tracking experiments. Through the statistical analysis of the quantitative characteristics of eye tracking indicators (such as fixation duration, saccade amplitude, scanning path length, and residence time), the strategy or cognitive load of the subjects when using maps to solve the task was obtained. In 2017, Luká Herman et al. [25] proposed a new tool (i.e., 3DgazeR) for analyzing the eye movement data of interactive 3D models that uses an eye movement tracker to obtain eye movement data, which are stored in a CSV table, and proposed and evaluated five visualization methods: 3D raw data, 3D scanning path (gaze and scan), 3D attention map (heat map), animation, and z-coordinate graph change with time. Eye movement experiments have been used in video research. In 2009, Vulal U et al. [26] proposed a concentration method for surveillance videos based on eye gaze. The eye tracking method is used to obtain the moving objects in surveillance videos to determine which moving objects are ignored by users. In 2012, Hadizadeh H et al. [27] proposed to construct an eye movement database of standard video sequences through eye movement experiments. In this study, the researchers collected original eye gaze data, directly output the data in a CSV table, and processed and marked the data manually as valid or invalid according to a standard.

III. Method

In this study, the influencing factors of fusion expression are analyzed firstly according to different video–geographic scene fusion expression modes, and the parameter model of the video–geographic scene fusion expression mode is constructed. Second, according to different video–geographic scene fusion expressions, the corresponding eye movement mode is constructed. Then, an eye movement experiment is performed, and the user's eye movement data are analyzed according to the eye movement mode. Finally, through analysis, the expression effect of geographic scene and video information fusion is obtained and evaluated. The specific technology roadmap of this study is shown in Fig. 2.

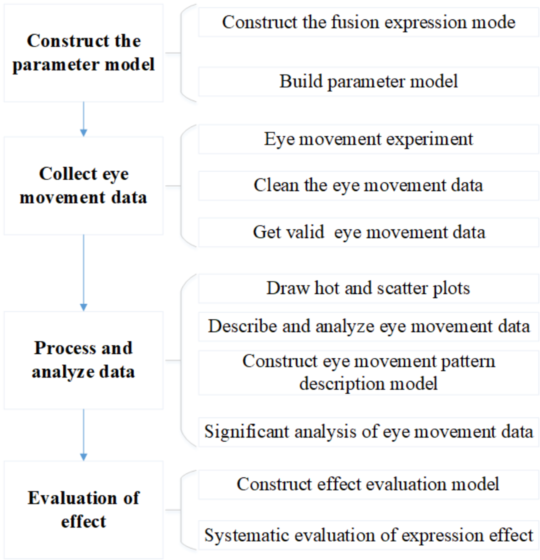


Fig. 2. Implementation process of this study.

A. *Fusion Expression Parameter Model*

In this paper, through the construction of a video-geographic scene fusion expression parameter model, the influencing factors of video–geographic scene fusion expression are comprehensively considered and classified; through the construction of a user parameter model in the eye movement experiment, the influence of individual differences, such as gender and age, on the video–geographic scene fusion expression effect is analyzed.

The parameter model of video–geographic scene fusion expression is constructed. First, the fusion expression mode of the video–geographic scene is constructed. The parameter model of video–geographic scene fusion expression is as follows:

$$\Phi\varsigma(A)= \{A1, A2, A3, ..., A10\}. \quad (1)$$

Among them, A1, A2, A3, ..., A10 represent the parameters that affect the fusion expression of the video-geographic scene. The parameters involved are A1 (geographical scene), A2 (video display rate), A3 (display object), A4 (fusion mode), A5 (embedding mode), A6 (number of video targets), A7 (video target projection fusion mode), and A8 (display size), A9 (video object projection fusion method), A10 (size of video or video object). Our model defines every value of each factor specifically, so we can quantify each factor.

According to the individual differences related to the experiment, the user parameter model of the eye movement experiment is constructed. The user characteristic parameter model is as follows:

$$\Phi\varsigma(B)= \{B1, B2, B3, ..., B6\}. \quad (2)$$

Among them, B1, B2, B3, …, B6 are the user characteristic parameters that affect the evaluation effect of fusion expression under the same conditions. The parameters are B1 (age), B2 (gender), B3 (occupation), B4 (education), B5 (whether have used surveillance video), and B6 (whether the uncorrected or corrected visual acuity is normal). B2 (gender) is the variable parameter.

B. Process and Analyze the Eye Movement Data

The specific processing and analysis of the eye movement experiment data are as follows:

First, we draw the hotspot and scatter maps of the eye movement data and intuitively analyze the fusion expression of the video–geographic scene under different influence factors. Second, the eye movement data is analyzed. Accuracy and precision are used to evaluate the three influencing factors (i.e., video compression, fusion expression of video–geographic scene and gender). In terms of accuracy evaluation, the first to evaluate is the accuracy of the subjects $R\tau$, which is the accuracy of completing the experimental task according to the movement direction of the target person filled in by the subjects. In terms of precision, precision $R\alpha$ is defined as the percentage of fixed points distributed in the specified AOI.

$$P\alpha = N\alpha / N\tau \cdot 100\% \quad (3)$$

where $N\alpha$ is the number of fixation points distributed in the specified AOI, and $N\tau$ is the total number of fixation points. Third, the eye movement pattern of the subjects watching the video is described, and the eye movement pattern description model is constructed. Finally, through the significant analysis of eye movement data, the impact of the factors selected on the performance of video–geographic scene fusion is determined.

IV. Design of Experiment

A. Participants

Eight testers, that is, four boys and four girls with ages between 22 and 25, participated in this experiment. All the testers had normal naked or corrected visual acuity, no color blindness or weakness, and good visual resolution. These testers are not eye tracking researchers but have a certain understanding of eye tracking technology.

B. Apparatus

In this experiment, a virtual reality helmet was used to simulate an eye tracker to obtain eye movement data. The HTC VIVE VR helmet used was produced by HTC and integrates steam VR tracking 1.0 technology and a partner guidance system, with a combined resolution of 2160 × 1200 (monocular resolution of 1080 × 1200), a 90 hz refresh rate, and a tracking accuracy of 0.1°. Steam VR was used in the software, which was installed in a computer equipped with an NVIDIA GTX1070 graphics card to drive the HTC VIVE VR device, and the unity2017.4.19f1 software was used to record the fixation data of the testers and analyze the visual fixation. The tester completed the test independently in a quiet environment without any help.

C. Materials

Our experiment was carried out in a 2D geographical scene, and the scene and video data were selected from the open source provided by DukeMTMC [28]. This experiment has three video-geographic fusion scenes which includes two video image-scene fusion and a video object-scene fusion. And the front two scenes are with three videos in the same location, and the last one is with almost all the video object in the three videos in the same location. This experiment is divided into two categories. In the first category, video images and objects are played at normal speed. In the second category, video images and objects are played according to the speed after video compression. Each category is further divided into three sub-categories. In the first sub-category, the surveillance video is embedded according to the camera position, and the testers look for the pedestrians with a blue coat. In the second sub-category, the surveillance video is connected with its position by line, and the testers look for the pedestrians with a red coat. In the third sub-category, the video data sub image is projected to the corresponding position, foreground projection is realized, and the testers look for the pedestrians with a white coat. The background image is shown in Fig. 3.

Fig. 3. Background of the experiment.

The experimental scenes of embedded, correlation, and foreground projection are shown in Fig. 4.

Fig. 4. Experiment scenes of embedded, correlated, and foreground projection.

D. Implementation

Before the experiment, we introduced the experiment to the testers. The testers were asked to sit at the origin of the coordinates. All testers underwent a pre-test to determine the best test location and ensure the accuracy of the test results. We divided the boys into two groups and the girls into two groups. One group of boys and girls did experiments one to three, and the other group of boys and girls did experiments four to six. According to the expression parameter model of video–geographic scene fusion, three typical video–geographic scene fusion modes were designed. The three fusion modes can be divided into video image + scene and video object + scene. The eye movement experiments were carried out by using the three video–geographic scene fusion expression modes as models. The three-video geographic scene fusion expression modes are as follows. (1) Video image + scene: ① Distributed embedded (video embedded in the geographical scene according to the angle of each camera); ② Correlation (in a geographic scene, it just points out the location and direction, and connects to the video image with a line). (2) Video object + scene: ③ Foreground projection (motion track + sub image of all video objects in a certain period, fusion display in the geographical scene). The specific task design is shown in Table I.

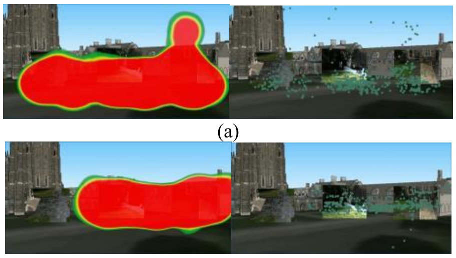
TABLE I. EXPERIMENTAL TASK DESIGN TABLE

Experimental Task Design Table					
Experiment	***Display objects***	***Video- geographic scene fusion expression mode***	***Video compressed***	***Groups***	***Task***
Experiment One	Video image	embedded	Uncompressed	First group of boys and girls	Pedestrian with blue coat
Experiment Two	Video image	Correlation			Pedestrian with red coat
Experiment Three	Video object	Foreground projection			Pedestrian with white coat
Experiment Four	Video image	embedded	compressed	Second group of boys and girls	Pedestrian with blue coat
Experiment Five	Video image	Correlation			Pedestrian with red coat
Experiment Six	Video object	Foreground projection			Pedestrian with white coat

V. ANALYSIS OF RESULTS

The experimental results are analyzed from two aspects: (1) the impact of video compression on video–geographic scene fusion expression for male and female subjects; (2) the impact of three different video–geographic scene fusion expression methods on video–geographic scene fusion expression for male and female subjects. Through the hotspot and scatter maps, the visual differences of different influencing factors are analyzed, and the eye movement data are analyzed using the significance test. The results can show whether differences exist among the influencing factors and whether the differences are significant to judge whether the influencing factors have an impact on the fusion expression effect.

A. *Influence of Video Compression on Video–Geographic Scene Fusion Expression*

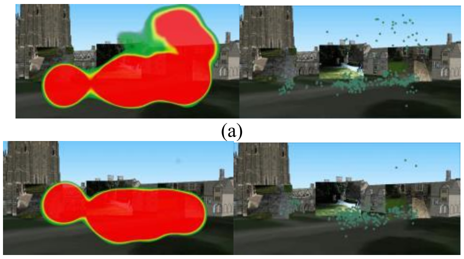


(a)

(b)

Fig. 5. Hotspot and scatter maps of uncompressed embedded video into geographical scene: (a) male; (b) female.

Fig. 5 (a) and (b) present the hotspot and scattered maps of the male and female subjects' task completion for the uncompressed video embedded in the geographical scene. The figures show that a part of the female eye movement data is missing after the experiment. Men's fixation is relatively scattered, and they pay more attention to the background environment than the women do. It is indicated that men and women have different visual fields.



(a)

(b)

Fig. 6. Hotspot and scatter maps of compressed embedded video into geographic scene: (a) male; (b) female.

Fig. 6 (a) and (b) respectively show the hotspot and scatter maps of the male and female subjects' task completion for the compressed video embedded in the geographical scene. According to the hotspot and scatter maps, the women remain focused on the task completion itself, mainly focusing on the location of the video, whereas the men are more concerned about the background after video compression.

1) *Accuracy and precision of task completion*

The statistics of the accuracy and precision of the task completion (for uncompressed and compressed videos) are shown in Table II.

TABLE II. ACCURACY AND PRECISION OF THE TASK

(%) ***Classification***	**Video compression**			
	Uncompressed		***Compressed***	
	male	female	male	female
Accuracy $R\tau$	92.59	100	83.33	94.44
Precision $R\alpha$	95.31	67.67	86.90	84.20

For the uncompressed video, the accuracy of the males' task completion is 92.59%, and that of the females' task completion is 100%. Thus, women are more accurate in completing tasks. In terms of precision, the precision of men is much higher than that of women due to the women's loss of eye movement data. For the compressed video, the accuracy of the males' task completion is 83.33%, and that of the females' task completion is 94.44%. Again, women are more accurate in completing tasks. In terms of precision, a slight difference exists between the men and the women. The comparisons in Fig. 5, Fig. 6, and Table II indicate that after video compression, the accuracy of the subjects in task completion decreased evidently, and the precision of the males decreased evidently.

2) *Significance Test*

Table III provides the *T*-test results of eye movement data for the uncompressed and compressed videos. The table shows that the significance of the *T*-test of the x- and y-coordinates is less than 0.05, and a significant difference exists between the uncompressed and compressed videos, indicating that the video compression has an impact on the fusion expression of the video–geographic scene.

TABLE III. INDEPENDENT SAMPLE T-TEST FOR UNCOMPRESSED VIDEO AND COMPRESSED VIDEO

Independent Sample *T* Test								
		F	***Significance***	***T***	***Free degree***	***Significance (two tailed)***	***Mean difference***	***Standard error difference***
x	Assumed equivariance	30.833	.000	-2.799	2986	.005	-.036425489	.013011910
	Equivariance is not assumed			-2.909	2861.429	**.004**	-.036425489	.012520600
y	Assumed equivariance	5.144	.023	9.290	2986	**.000**	.037598274	.004047376
	Equivariance is not assumed			9.298	2571.419	.000	.037598274	.004043845

B. *Influence of Video–Geographic Scene Fusion Methods on Video–Geographic Scene Fusion Expression*

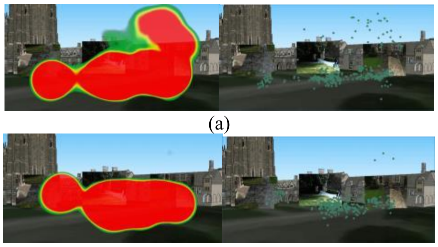

(a)
(b)

Fig. 7. Hotspot and scatter maps of compressed embedded video into geographic scene: (a) male; (b) female.

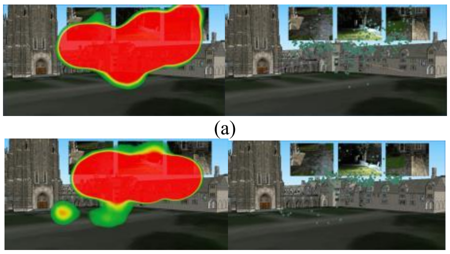

(a)
(b)

Fig. 8. Hotspot and scatter maps of compressed correlational fusion of video and geographic scene: (a) male; (b) female.

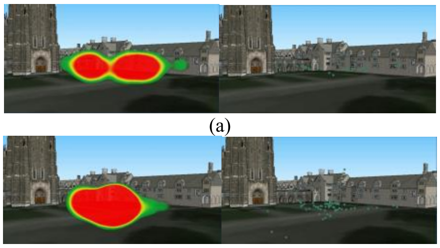

(a)
(b)

Fig. 9. Hotspot and scatter maps of compressed foreground projection fusion of video and geographic scene: (a) male; (b) female.

As shown in Fig. 7 (a) and (b), the compressed video is embedded into the hotspot and scatter maps of the geographical scene. The figure indicates that men also pay attention to the background, that is, the geographical scene, while girls pay more attention to the task itself. Fig. 8 presents the hotspot and scatter maps of the male and female subjects' task completion after video compression and geographic scene correlation fusion. No obvious difference exists between (a) and (b) in Fig. 8. Fig. 9 presents the hotspot and scatter maps of the male and female subjects' task completion after video compression and geographic scene foreground projection fusion. The comparison of the figures indicates that men's eye gaze is relatively scattered, while women's gaze is relatively concentrated. Given that this is the third experiment for each subject in the same background picture, the observation of the environment should be evidently reduced.

1) *Accuracy and precision of task completion*

The accuracy and precision of the subjects completing the task of video–geographic scene fusion expression are shown in Table IV. The table shows that in the embedded scene, the accuracy of men is evidently lower than that of women, while the precision of men is relatively higher. The accuracy and precision of the correlation fusion in the table indicate that the accuracy of men is less than that of women, and the precision of men's eye gaze is higher than that of women. Comparing with the accuracy of embedded fusion, the accuracy of correlation fusion is greatly improved, and the precision does not change much. For foreground projection, the accuracy is 100%, and the precision of women is higher than that of men. The accuracy of foreground projection fusion is higher than that of the former two fusion methods, and the precision does not change much. The accuracy of foreground projection is the highest among the three video–geographic scene fusion methods; women have higher accuracy in completing tasks than men.

TABLE IV. ACCURACY AND PRECISION OF THE TASK

(%)	**Video–geographic scene fusion method**					
Classification	***Embedded fusion***		***Correlation fusion***		***Foreground projection***	
	male	female	male	female	male	female
Accuracy $R\tau$	73.33	88.89	93.33	100	100	100
Precision $R\alpha$	96.71	82.62	93.78	78.00	75.00	89.33

2) Significance test

Table V shows the *F*-significance test of the three video–geographic scene fusion methods. Table V (a) indicates that the significance of the x-coordinate is 0.008 < 0.05, and the significance of the y-coordinate is 0.527 > 0.05. Thus, the *F*-significance test cannot be analyzed in the x-coordinate, and the significance of the y-coordinate is analyzed later. Table V (b) shows that the significance is 0.00, indicating significant differences among the three fusion methods of embedded, correlation, and foreground projection expression. Table V (c) indicates that a significant difference exists between the embedded and correlation expressions, no significant difference exists between the embedded and foreground projection expressions, and a significant difference exists between the correlation and foreground projection expressions. Table VI presents the *T*-test results on the gender factors of the eye movement data. The table shows that the significance of the *T*-test of the x-coordinate is 0.006 < 0.05, and that of the *T*-test of the y-coordinate is 0.62 > 0.05. Therefore, significant differences in gender factors exist, indicating that gender has a certain influence on the fusion expression of the video–geographic scene.

TABLE V. F-SIGNIFICANCE TEST OF DIFFERENT FUSION METHODS

(a) Homogeneity test of variance

Homogeneity test of variance					
		Levin statistics	*Free degree 1*	*Free degree 2*	*Significance*
x	Based on average	4.813	2	1619	.008
	Based on median	4.233	2	1619	.015
	Based on median with adjusted degrees of freedom	4.233	2	1576.922	.015
	Based on the average after clipping	4.886	2	1619	**.008**
y	Based on average	1.357	2	1619	.258
	Based on median	.605	2	1619	.546
	Based on median with adjusted degrees of freedom	.605	2	1543.633	.546
	Based on the average after clipping	.641	2	1619	**.527**

(b)ANOVA

ANOVA						
		Quadratic sum	*Free degree*	*Mean square*	*F*	*Significance*
y	Interblock	29.763	2	14.881	1560.663	.000
	Intragroup	15.438	1619	.010		
	Aggregate	45.201	1621			

(c) Multiple comparisons

Multiple comparisons							
dependent variable	*(I) Fusion method*	*(J) Fusion method*	*Mean difference (I-J)*	*Standard error*	*Significance*	*95% confidence interval Lower limit*	*upper limit*
y	Embedded fusion	Correlation fusion	-.275651938*	.005044128	**.000**	-.28554564	-.26575823
		Foreground projection	-.017052961	.009766594	**.081**	-.03620945	.00210353
	Correlation fusion	Embedded fusion	.275651938*	.005044128	**.000**	.26575823	.28554564
		Foreground projection	.258598977*	.009866283	**.000**	.23924695	.27795100
	Foreground projection	Embedded fusion	.017052961	.009766594	**.081**	-.00210353	.03620945
		Correlation fusion	-.258598977*	.009866283	**.000**	-.27795100	-.23924695

TABLE VI. INDEPENDENT SAMPLE T-TEST OF GENDER FACTORS

Independent sample *T*-test								
		F	*Significance*	*T*	*Free degree*	*Significance (two tailed)*	*Mean difference*	*Standard error difference*
x	Assumed equivariance	46.979	.000	2.744	810	.006	.056329510	.020524801
	Equivariance is not assumed			2.743	756.239	**.006**	.056329510	.020538060
y	Assumed equivariance	55.055	.000	-.304	810	.762	-.002218451	.007309457
	Equivariance is not assumed			-.303	572.994	**.762**	-.002218451	.007320988

VI. CONCLUSION

The proposed method combines virtual reality technology with eye tracking theory, uses a VR helmet for eye tracking, and realizes video–geographic scene fusion expression. The experimental results show that for compressed and uncompressed videos, the video–geographic scene fusion expression method and the gender of the testers have a certain impact on the video–geographic scene fusion expression effect. The accuracy of watching an uncompressed video is higher than that of watching a compressed video. In terms of the video–geographic scene fusion expression, no obvious difference exists between the embedded and the foreground projection. However, an obvious difference exists between the correlation expression and the previous two expressions. In terms of viewer features, the accuracy of the male's task completion is lower than that of the female's task completion. This study has a certain reference value for the research and development of video–geographic scene fusion expression.

VII. Limitations and Future Work

This paper analyzes the influence of the video compression state, the video–scene fusion expression mode, and the viewer characteristics on the video–geographic scene fusion expression, and there are still many factors to be analyzed. In this study, the fusion expression of the 2D geographic scene and video is designed. In the future, the fusion of video and the 3D scene can be used to analyze the expression effect. In this experiment, few researchers were invited, and the data is insufficient. Thus, the results of this experiment may be accidental. Increasing the amount of data for analysis can be considered in the future.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (NSFC) (Grant numbers 41771420 and 41801305), the Project of Natural Science Research in Colleges and Universities in Jiangsu Province (Grant number 18KJB170007), the Postgraduate Research and Practice Innovation Program of Jiangsu Province (Grant numbers KYCX20-1323 and KYCX20-1180), and the State Scholarship Fund from the China Scholarship Council (CSC) (Grant number 202006860047).

References

[1] H Brett and T Mohan, "A novel graphical interface and context aware map for incident detection and monitoring," 9th World Congress on Intelligent Transport Systems, 2002.

[2] EG Rieffel, A Girgensohn, D Kimber, T Chen, and Q Liu, "Geometric Tools for Multicamera Surveillance Systems," 2007 1st ACM/IEEE International Conference on Distributed Smart Cameras, ICDSC. pp.132-139, 2007.

[3] Y. Y. Chen, Y. H. Huang, Y. C. Cheng, and Y. S. Chen, "Integration of multiple views for a 3-d indoor surveillance system," International Journal on Information, vol.13(6), pp. 2039-2057, 2010.

[4] R Du, B Sujal, and V Amitabh, "Video fields: fusing multiple surveillance videos into a dynamic virtual environment," The 21st International Conference. ACM, 2016.

[5] A Lippman, "Movie maps: An application of the optical videodisc to computer graphics," SIGGRAPH'80, vol.14(3), pp.32-43, 1980.

[6] J K Berry, "capture 'where' and 'when' on video-based GIS ," GE0W0RLD, vol.13(9), pp.26-27, 2000.

[7] S Y Lee, S B Kim, and J H Choi, "4S-Van: A prototype mobile mapping system for GIS," Korean Journal of Remote Sensing, vol.19(1) pp.91-97, 2003.

[8] T H Hwang, K H Choi, I H Joo, and J H Lee, "MPEG-7 metadata for video-based GIS applications," IGARSS'03 Proceedings, pp.3641-3643, 2003.

[9] S. Y. Lee, K. H. Choi, I. H. Joo, S. I. Cho, and J. H. Park, "Design and implementation of 4s-van: a mobile mapping system," ETRI Journal, vol.28(3), pp.265-274, June 2006.

[10] T Navarrete and J Blat, "VideoGIS: Segmenting and indexing video based on geographic information," 5th AGILE conference on Geographic Information Science, Spain: Palma de Mallorca, pp. 1-9, 2002.

[11] N Pissinou, I Radev, and K Makk, "Spatio-temporal modeling in video and multimedia geographic information systems," GeoInfomatica, vol. 5(4), pp. 375-409, 2001.

[12] I H Joo, T H Hwang, and K H Choi, "Generation of video metadata supporting Video—GIs integration," IC1P'04 Proceedings, Singapore, pp.1695-1698, 2004.

[13] B. Tang and M. Y. Zhou, "GIS of existing lines based on video image," Railway Computer Application, vol.10(11), pp.31-33, 2001.

[14] Y. F. Kong, "Design and implementation of a highway video GIS," Highway, pp.119-121, 2007.

[15] Y. F. Li and J. L. Zhu, "Design and development of railway line video data acquisition system," Railway Computer Application, vol.13(12), pp.4-6, 2004.

[16] J. F. Feng, H. Zhang, and Y. J. Sha, "Design of GPS vehicle mobile video monitoring system," Bulletin of Surveying and Mapping, vol.2, pp.52-54 2007.

[17] Y. Wu, X. J. Liu, H. Zhao, and Y Wu, "Research on video capture method with location," Bulletin of Surveying and Mapping, vol.1, pp.24-27, 2010.

[18] Y. F. Kong, "Design of geographic video data model and implementation of network video GIS," Geomatics and Information Science of Wuhan University, vol.35(2), pp.133-137, 2010.

[19] Y. F. Kong, "Design and application of geographic hypermedia system based on Web Services," Journal of Geo-Information Science , vol.12(1), pp.76—82, 2010.

[20] H. Q. Song and Y. F. Kong, "Design and development of video GIS system in Adobe flex framework," Geomatics and Information Science of Wuhan University, vol.35(6), pp.743-746, 2010.

[21] Rayner and Keith, "Eye movements in reading and information processing: 20 years of research," Psychological Bulletin, vol.85(3), pp.618-660, 1978.

[22] W. H. Dong, H. Liao, X. U. Fang, Z. Liu and S. B. Zhang, "Using eye tracking to evaluate the usability of animated maps," Science China Earth Sciences, vol.57(003), pp.512-522, 2014.

[23] A Brychtova and A Coltekin, "An Empirical User Study for Measuring the Influence of Colour Distance and Font Size in Map Reading Using Eye Tracking." Cartographic Journal, vol.53(3), pp. 202-212, 2016.

[24] Popelka, Stanislav, and A. Brychtova, "Eye-tracking Study on Different Perception of 2D and 3D Terrain Visualisation," Cartographic Journal, vol.50(3), pp. 240-246, 2013.

[25] Luká H, S. Popelka, and Vendula H, "Eye-tracking Analysis of Interactive 3D Geovisualization," Journal of Eye Movement Research, vol.10(3), 2017.

[26] U Vural, and Y S Akgul, "Eye-gaze based real-time surveillance video synopsis," Pattern Recognition Letters, vol.30(12), pp.1151-1159, 2009.

[27] H Hadizadeh, M J Enriquez, and I V Bajic, "Eye-Tracking Database for a Set of Standard Video Sequences," IEEE Transactions on Image Processing, vol.21(2), pp.898-903, 2012.

[28] http://vision.cs.duke.edu/DukeMTMC/