



Computer-Assisted Human-Computer Interaction in Visual Communication

Ying Zhang 

School of Arts, Henan University of Animal Husbandry Economics, Zhengzhou, Henan 450000, China, smile0701z@163.com

Corresponding author: Ying Zhang, smile0701z@163.com

Abstract. With the help of science and technology, digital media expressions continue to innovate, which has entered a new era and driven the development of computer-computer interaction technology. In the field of visual interaction art design and related technologies, the development of visual interaction and related field applications are still less studied. This article will explore the development process, application fields and related technologies of visual interaction from the perspective of technicians. In combination with the development of digital media art, we discussed how to better combine the visual interaction experience with creative design, and analyzed the key points of the visual interactive work creation process, design, and technology. In addition, this article will build an image processing model for visual communication of human-computer interaction interface, use edge contour feature extraction method to extract its boundary information features, combine cross-information fusion technology to achieve enhanced processing of its visual communication information, and construct an edge package for its visual communication image. The network feature realizes the optimized design of the visual communication of the human-computer interaction interface. The test results show that this method can significantly improve the fidelity of its visual communication, improve the discrimination ability of the interface output, and effectively help show the communication effect of visual interactive works.

Keywords: Visual communication; human-computer interaction; Visual transmission; Computer-assisted communication

DOI: <https://doi.org/10.14733/cadaps.2021.S1.109-119>

1 INTRODUCTION

In the digital age, with the emergence of "off-desktop" technologies such as virtual environments, augmented reality, mixed reality, and computer interaction, the creation and development of multi-channel human-computer interaction design art has become an important research direction. Katie Li et al. [1] proposed that if an inappropriate visualisation method is employed it could hinder the design process. Computer interaction is particularly valued in the art and science

communities by Ackermann et al. [2] At present, the technology supporting computer vision is gradually maturing, and non-traditional vision devices and interface components are rapidly increasing and becoming more common. In the fields of digital media, art display, intelligent visual monitoring, science and information visualization, games and entertainment, visual technology is also increasingly used, which requires a well-designed user interface. Barnes et al. [3] proposed that visual interaction has become one of the hottest directions in human-computer interaction (HCI) research today, and it is the frontier of innovation in multidisciplinary fields.

"Visual" human-computer interaction refers to the principle of obtaining information from human vision, using a computer in conjunction with a camera, and detecting and identifying user operations using non-traditional input devices, thereby realizing natural human-computer interaction in two or three dimensions. Chakraborty et al [4] proposed that in this process, the camera captures the physical target in the environment, converts the obtained image into a digital signal and transmits it to the computer. The computer realizes the processing and understanding of "visual" information, which includes detecting and analyzing the position and motion state of the target for interactive control. Cheng K. et al. [5] proposed that human-computer interaction based on "vision" is one of the communication modes between the human-machine interface and the user. Hallgreen et al. [6] think that the identified visual displays were grouped into visual types and users must communicate instructions, needs, questions, intents, and goals to the system, and accordingly, the system must provide feedback, input requests, system status information, and so on. We can think of this communication process as a series of conversions. The user and the system do not use the same language, so the interface must be able to act as a translator. In fact, it contains multiple translation steps (see Figure 1); the user first converts the target into actions, then the input device converts these physical actions into a set of electronic forms that the system can accept, and finally the system interprets them according to the current system status these signals. Hossan et al. [7] introduced a human bond communication (HBC) method, based on head-mounted displays (HMDs), which enables continuous bidirectional communication among multiple users with or without Internet connections. Usually the system responds to the user's input in some way, so there must be some conversion between the input and output-called the transfer function. In order to convey the output to the user, the system must convert the information into a digital display representation, and then through the output device into a visual perception signal that the user can receive. Finally, the user translates these perceptions into meaningful semantic representations. The conversion between input and output is known as the control-display mapping produced by Lee H. Kim et al [8].

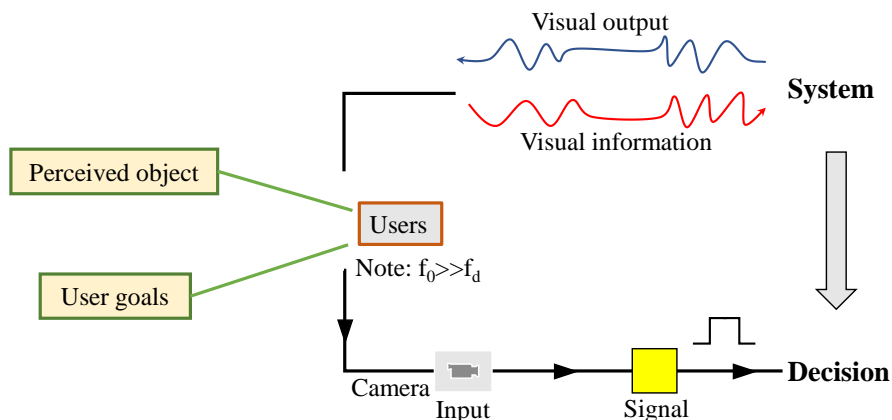


Figure 1: Multiple translation steps in "Visual" human-computer interaction.

Under the application of the current computer science new technology innovation, the information conveyed by the interactive expression form can be separated from the traditional human-

computer interaction, and also provide new innovative thinking for designers in different fields of human-computer interaction. As Lou, Y. et al. [9] mentioned, the essential building blocks of the unified visual data management and analysis infrastructure are introduced to facilitate the utilization of the visual signal in the artificial intelligence era. This article mainly focusing on the development direction based on the "visual" human-computer interaction, application field, the design creation and related technology application perspective to study the visual interaction brings the user experience and feelings, and refer to the principle of human visual access to information. At the end of the paper used a computer with a USB camera operations research of target detection and motion analysis, we also realize the nature of two-dimensional or three-dimensional space process of human-computer interaction. In this process, the camera captures a physical object in the environment and converts the resulting image into a digital signal to the computer. This paper by introducing the history of visual interaction and the future development direction and analyzing digital media interaction design creation method and related technology research to explore how the visual interactive form appear more directly, users get good interactive experience in visual interaction, which can give the user more interesting and a good show, thus extending the best human-computer interaction effect.

2 REALIZATION OF KEY TECHNOLOGY OF CONNECTING ROD MOLD

2.1 Related Work in the Field of Visual Interaction Applications

In the field of digital media design, visual interaction can express more meta-visual shapes, which affects the creative thinking. The creation of image shapes is no longer limited to the author's own description and records. The image of the human body in the exhibition space is captured and analyzed by the camera. The dialogue channel between viewers and works was opened. During the interaction, subtle combinations of dynamic image elements are produced. As the timeline progresses, various unpredictable visual effects change. Over the past ten years, some results have gradually developed. Especially in the field of interactive design, it has attracted the attention of design creators in the past five years, and developed various real-time interactive media art designs through several different imaging technologies. Here are some common expressions:

(1) Tanić M et al. [10] use computer vision to detect whether objects in the frame are in motion. Authors capture the position of the image at the moment of motion to produce a dynamic visual effect. *Playing-with-fire* uses this concept to create, depicting the effect of flame burning on the corresponding position of the image between the limb movements, and produces a rich visual modeling effect with the frequency and amplitude of the limb swing;

(2) Wu H. et al. [11] obtain the human outline in the picture through computer vision. This method can mainly design the effect of virtual objects such as sticking to the body. The work *Molecular Bubbles* uses this concept to create, using projection to place it behind the viewer. *Shadows* are created on the screen due to backlight illumination. When the virtual object in the screen touches the silhouette of a person, it will produce a special visual effect. In this work, the shape and appearance of the virtual object will be changed due to human intervention, resulting in interesting interaction situation;

(3) Authors through some methods for object tracking, from simple single object to multiple object tracking, obtain the position of the object and even obtain the outline. The work *The Manual Input Sessions* presents creations in the form of live performances, and uses flexible mastery of sound and video to create a clear-cut, humorous, interactive sound art. Through the computer program to analyze the shape of the object, and automatically calculate the unique sound based on the contour, it can also recognize the shape of the hand, and produce a unique sound effect based on the posture of the palm, creating a rich listening experience.

2.2 Entertainment Games

In the field of entertainment games, visual interactive games have entered a stage of rapid development. It can interact with the game by capturing user action information, as shown in Figure 2. These games are getting more and more attention and praise from players and the game industry. This type of non-traditional advanced input control method has been included in the range of new generation game console standard input devices. One of the most famous is Sony's EyeToy, which was launched by Sony in 2003. The USB camera bundled with ps2 controls the input game. It can interact with the game by capturing user action information. EyeToy did not achieve great success in the market in the early days. Due to many factors such as technology and player habits, EyeToy games are always limited to some "childlike" games in design. SCEA of Sony North America has obtained the technology license of GestureTek's Video-Gesture-Control, which specializes in the development of posture control technology and holds a number of patents. Now EyeToy's game types are becoming more and more diversified, becoming more and more fun, and in recent years have gradually attracted the attention and praise of gamers and the game industry. This type of non-traditional advanced input control method is considered in the standard input device of the next generation game console. Sony has invested huge manpower and material resources in the research of visual interaction, and these are not just to provide some auxiliary games on the PS platform. The most fundamental purpose is to promote the advancement of visual interaction technology. The efficiency of input based on visual interaction technology is much higher than traditional input methods. Traditional input methods can lose a lot of useful data. For example, the "Contra" game that was popular on the red and white machines in the 1980s. When playing the game, when the character needs to jump up, his hands holding the handles are always unconsciously swinging up. This typical example reflects the most natural human-computer interaction phenomenon. We need to interact with the computer in a more natural way. The emergence of EyeToy is exactly the possibility provided for realizing the long-awaited game mode. Visual interaction has a very broad application prospect, and its application in the field of digital entertainment will bring users a more natural gaming experience.

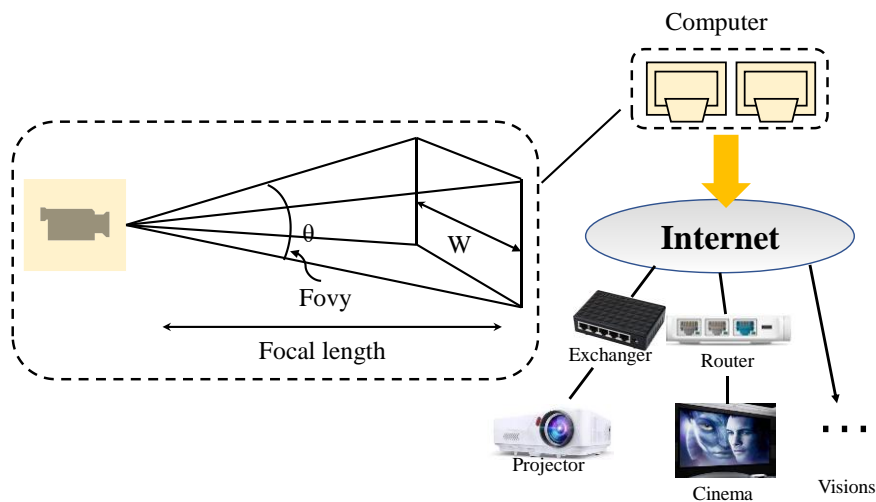


Figure 2: The main process in the construction of visual interactive games.

2.3 Visual Surveillance

In the field of intelligent visual surveillance, the method of computer vision is used to locate, identify and track targets in dynamic scenes by automatically analyzing the image sequence recorded by the camera with almost no human intervention. Based on the analysis and judgment

of the target's behavior, it can not only complete the daily management, but also respond in time when abnormal situations occur. Intelligent visual monitoring technology has a wide range of application prospects, and can be used in traffic scenarios, such as monitoring of intersections, highways, parking lots, and airfields, monitoring of military scenarios, monitoring of important national security agencies such as military bases, banks, and sensitive public area. Providing technical and methodological support for intelligent monitoring applications in future social security places is of strategic importance.

3 Computer-Aided Man-Machine Interaction

Human-computer interaction (HCI) is one of the most important application areas of computer vision. Currently, there are multiple modes of HCI, such as gestures, gaze directions, and voice. In these modalities, gesture actions are intuitive modalities for performing HCI. However, because of its low complexity and easy recognition, gesture pointing can be considered more natural. One of the man-machine interface based on speech recognition is a natural and flexible implementation method because it is most commonly used for human communication. However, in a noisy interactive environment, HCI based on speech is prone to cause interference, so its application often requires specific environmental requirements. With the development of ubiquitous computing, people need to emphasize the diversity and nature of human-computer interaction. In this paper, a new multimodal fusion method is proposed for real-time HCI in the equipment manufacturing industry. The main research contents are as follows: First, by integrating different interaction modes including gestures, gaze directions, and speech in a non-contact and non-wearable manner to determine the interaction target, and using human communication and interaction capabilities to develop a novel interaction method. This method eliminates some existing constraints that may exist in real HCI. In addition, interactive users can provide timely feedback on audiovisual integration through immersive experiences. Secondly, a multi-modal fusion method based on decision matrix is proposed to fuse these three modalities together to improve the convenience and accuracy of HCI.

3.1 Assembly Effect Comparison

HCI is significantly influenced by the communication tools that humans use to control computers. Gaze-driven controls have replaced some keyboard or mouse functions, opening up new depths of HCI. The user's gaze is more informative than moving the mouse, that is, not only an emotional factor, but also the direction of the gaze. In order to facilitate the research, the face gaze direction is applied to perform the HCI action conveniently and intuitively. The industry has proposed Active Shape Models (ASM) to achieve HCI analysis of the orientation of the face. In order to obtain the gaze direction, this paper uses depth maps and depth cameras Corresponding color images, the position and size of the face area are used to initialize the ASM model parameters, and the geometric distance between each two facial contour feature points is obtained to improve the stability of gaze direction estimation. N facial contour feature points are extracted in the ASM color image, and three key facial feature points are used to construct the following 2D facial geometric model. After processing through the calibrated depth camera, the 2D facial geometric model is mapped to the 3D coordinate space. For example, two eye feature points and a nasal tip point are selected to construct an initial triangular surface as a judgment of the gaze direction, which is shown as follows:

$$n_e = (X_e, Y_e, Z_e) = P_A P_B \otimes P_A P_C \quad (1)$$

$$P_A P_B \otimes P_A P_C = (X_A - X_B, Y_A - Y_B, Z_A - Z_B) \otimes (X_A - X_C, Y_A - Y_C, Z_A - Z_C) \quad (2)$$

Where, \otimes means cross operator, PA and PB mean the eye features, PC means nose features.

The gaze direction-based interaction point is structured to have an origin for HCI, expressed as a formula:

$$T_{et} = (X_{et}, Y_{et}, Z_{et}) = P_{eye_mid} + c_0 n_e \quad (3)$$

Where, c_0 is the zero scalar of Z_{et} .

Because the depth map has some defects, including occlusion areas and small black holes, there are some random incorrect depth data. In addition, the position of the facial contour feature points fluctuates with the rotation of the face, and since the facial interaction points are different from each other, it is difficult to perform stably HCI. To solve this problem, the time median filter strategy is developed as follows, using several consecutive frames to execute the time median filter to determine the interaction point of the gaze direction, as shown in formula (4):

$$T_e = \frac{1}{F_e} \sum_{i=0}^{F_e-1} T_{et}(i) \quad (4)$$

Where $T_{et(i)}$ is the coordinate of the interaction point in frame i . F_e is the frame number selected for the time median filter.

During the experiment, we found that the performance of HCI is mainly determined by a series of continuous video frames (F_e and F_p). Larger values of F_e and F_p increase processing time, while smaller values lead to erroneous interaction results. In order to make a reasonable trade-off between efficiency and interaction performance, the variance of the inspection target is used to measure the stability of gaze and finger-based HCI systems. The variance varies with different frames. The experimental results are shown in Figure 3.

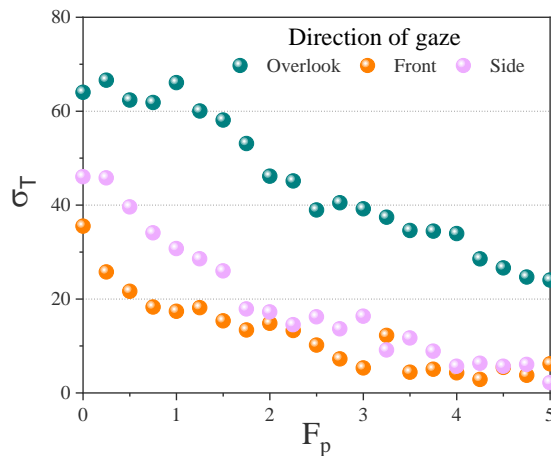


Figure 3: The F_p related change of T in different direction of gaze.

3.2 Gesture Direction Estimation

Gesture is another evaluation method designed by HCI. It does not require any prior skills or training. When a user performs a finger-based HCI, his pointer is closer to the target than the other hand. In order to distinguish the interactive pointer, the 3D coordinates pointing to the hand joint can be obtained from the skeleton diagram by formula (5):

$$P_h = \begin{cases} P_r & \text{if } Z_r < Z_i \\ P_i & \text{else} \end{cases} \quad (5)$$

Among them, P_r and P_i represent the right and left hands of the recognition object, and Z_r and Z_i represent the Z coordinates of the right and left hands, respectively. To perform HCI manual pointing, separate the hand from the depth map. By detecting and tracking the joints of the hands and wrists in the skeletal joint map, the pointing direction is obtained. The depth of the hand joint

can be calculated by the skeleton-to-depth conversion. The threshold value is used to apply the hand region $D_h(i, j)$ is separate from the original depth image.

$$D_h(i, j) = \begin{cases} 1 & \text{if } d(i, j) < D_p + \varepsilon_0 \\ 0 & \text{else} \end{cases} \quad (6)$$

Among them, $d(i, j)$ belong to D , m and n means the width and height. The process of determining the interaction target through the pointer can be divided into three phases namely: pointing, moving and non-gesture. The tracking of the pointing fingertip is as follows: In order to avoid pointing to the wrong interactive target, a Kalman filter strategy is used to record the detected trajectory of the pointing fingertip. The movement direction of the pointer is based on the 3D coordinates of the hand and wrist joints taken by the depth camera Determined, as shown in formula (7):

$$P_w P_{ft} = (x_{ft} - x_w, y_{ft} - y_w, z_{ft} - z_w) \quad (7)$$

Among them, the P_{ft} and P_w means the positions of fingers and wrist, respectively. In order to realize the visual communication design of the human-computer interaction interface based on hierarchical analysis, a 4×4 block combination model was used to sample the visual characteristics of the human-computer interaction interface. The target template matching matrix $\{w_1, w_2, \dots, w_i\}$ represents the vector weighting of the visual communication image extraction of the human-computer interaction interface, and the adaptive contour analysis method is used to calculate the edge contour feature set of the human-computer interaction interface.

$$\begin{cases} w_1 = P_{ft} - P_w \\ w_2 = P_2 - P_{ft} \\ w_3 = P_3 - P_2 \\ w_4 = P_w \end{cases} \quad (8)$$

3.3 Audio Speech Recognition Estimation

Informative speech is considered one of the indispensable forms of communication, which is the intuitive form of HCI. A key issue is how to perform HCI in audio information, as shown in Figure 4 depicting the process of audio speech recognition.

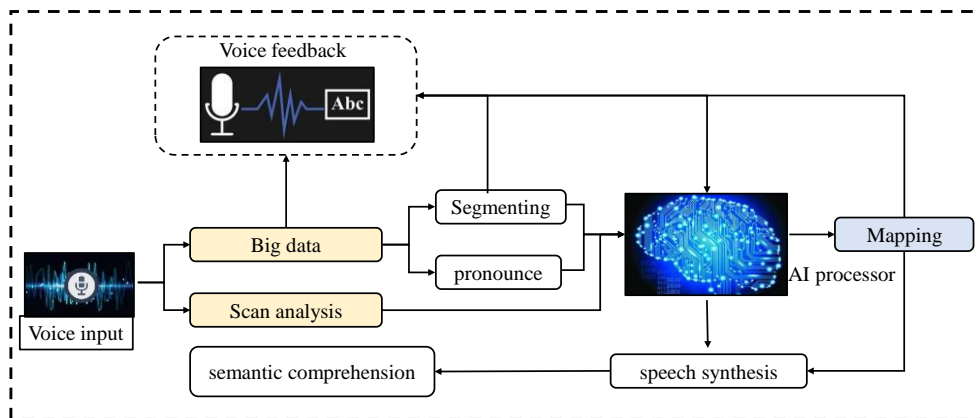


Figure 4: The process of audio speech recognition.

4 Computer Aided Visual Interaction Design and Implementation

Visual interaction greatly affects works created by computers. It enables creators and other developers to experience the interactive effects of computer applications together. It can also use technology to convey what the creator really tells. "Visual Interaction Design" incorporates aesthetics, humanities and technology. Visual interaction design is the use of vision to define the relationship between items; that is, to discuss how people use the senses, physical movements, etc. to interact with the environment, and more importantly, the focus of "visual interaction design" is on the quality of interaction and Reasonably. From the HCI mode, it is more direct and natural to replace the mouse and keyboard with an interactive physical interface than to control the graphics through the mouse and keyboard. Figure 5 indicates that the user an interactive interface with the computer media is created by the physical interface.

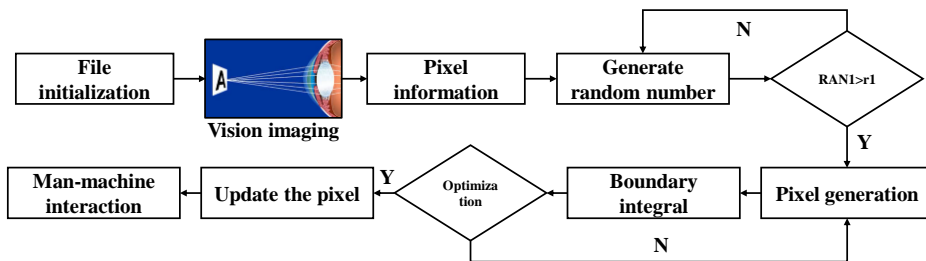


Figure 5: Computer Aided Visual Interaction Design and Implementation.

Laurel proposed that HCI is a design for enjoyment and experience in the research category of human-computer interaction. All interfaces should be as close to human, natural and direct experience considerations as possible. Consider design in terms of direct, natural, and real-time interaction. The most natural way of interaction should be in the form of direct manipulation and direct intervention. Later, others put forward a more specific visual human-computer interaction process. They explained how to use more intuitive ways to interact with visual senses and limbs. The computer communicates with each other, and gives instructions through actions, which are passed through the computer's visual dynamic tracking technology, so that the computer recognizes the instructions and information sent by the user, so as to provide visual feedback and achieve real-time visual interaction effects, as shown in Figure 6.

In addition, in visual interactive applications, many new algorithms for face detection have appeared, such as methods based on color information, contour detection, and gray distribution based on faces. Since the skin color of the face is still the most important and obvious feature of the face, the detection method based on color information uses an appropriate skin color system and uses the skin color as the main parameter for face detection. With the advantages of insensitivity and reduction and small deformation, the algorithm is easy to analyze the face area of the next frame of image based on the analysis result of the previous frame. In addition, it has easy analysis and calculation, which is more suitable for real-time systems. Then the face is tracked with the "Conditions Probability Density Propagation Algorithm". The random sample set is used to represent the probability distribution of the tracking target in the image. At the same time, in each iteration step, condensation generates a large number of random samples near the predicted position, which enhances its robustness to irregular motion. The steps for implementing face detection and tracking based on the above method are detailed below. The first step uses the characteristics of the HIS color system of the human skin color. Through filtering, we can obtain the human face area. Although the distribution of various faces in the HIS color system is equivalent, they are different after all.

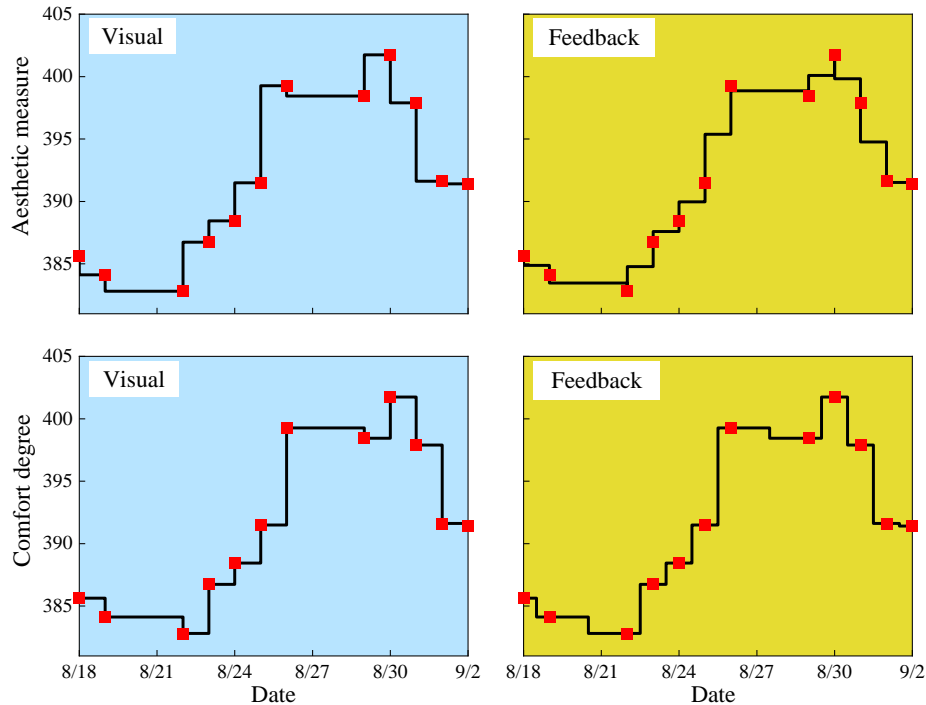


Figure 6: Visual feedback and achieved real-time visual interaction effects.

Besides, in order to make the algorithm more versatile, the distribution maps of the faces are randomly calculated and averaged, thereby obtaining a more general sense. And we use this histogram as the projected standard histogram. In this way, using the same histogram can save the process of calculating the histogram each time, thereby improving the overall detection speed. We also optimize the visual communication of the human-computer interaction interface and VC environment, and give the human-computer interaction interface visual communication image sample size of 1200. The human-computer interaction interface visual communication feature information scanning frequency is 24kHz, and the spatial distribution of interface elements can be seen in Figure 7 for position parameters. It can be seen that the method in this paper can effectively realize the optimal design of visual communication of the human-computer interaction interface and test the output pixel enhancement results.

Moreover, interaction design should be based on usability considerations such as "easy to use" and "comfortable", and even from interaction with products, to interaction factors generated by people to communicate with them, to deeper sensory experiences, called "interactive beauty". In this study, we combined the interactive aesthetics with the statistics of the experimenters to quantify the degree of interactive aesthetics through negative feedback. As shown in the Figure 8 below, this method can improve the fidelity of the visual communication of the human-computer interaction interface, and the resolution of the interface output is better.

5 CONCLUSIONS

This article mainly discusses the art design, related technologies, and development history of vision-based interactive fields, application directions, and explores its development prospects. And describe a visual interactive presentation system that can replace traditional computer input

devices. The main purpose is to explore how to use visual interaction techniques to combine visual image display design with each other.

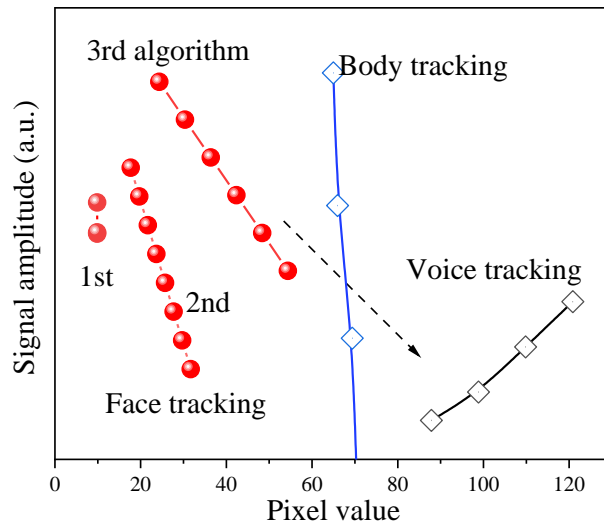


Figure 7: The spatial distribution of interface elements in different parts tracking.

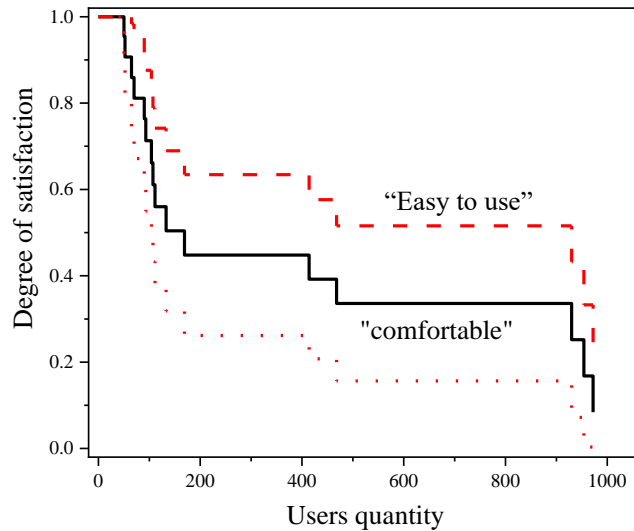


Figure 8: Quantified degree of interactive aesthetics through negative feedback.

Here, we also communicate and interact with the camera in the most natural way to get feedback, and use computer vision to interactively detect and analyze human behavior patterns in the scene. And they can continue to enhance the image effect and interaction mode in the future, which also show special visual effects. The research and analysis of this article prove that the computer-assisted human-computer interaction method restores the natural human interaction method, and

has revolutionized the form of interaction. This method could help us look forward to more interesting, imaginative and realistic creations in the future.

Ying Zhang, <https://orcid.org/0000-0003-1392-3114>

REFERENCES

- [1] Li, K.; Tiwari, A.; Alcock, J.; Bermell-Garcia, P.: Categorisation of visualisation methods to support the design of Human-Computer Interaction Systems, *Applied ergonomics*, 55, 2016, 85-107. <https://doi.org/10.1016/j.apergo.2016.01.009>
- [2] Ackermann, C.; Beggiato, M.; Schubert, S.; Krems, J.-F.: An experimental study to investigate design and assessment criteria: What is important for communication between pedestrians and automated vehicles? *Applied Ergonomics*, 75, 2019, 272-282. <https://doi.org/10.1016/j.apergo.2018.11.002>
- [3] Barnes, S.-R.: Appearance and explanation: advancements in the evaluation of journalistic information graphics, *Journal of Visual Literacy*, 35(3), 2016, 167-186. <https://doi.org/10.1080/1051144X.2016.1278109>
- [4] Chakraborty, B.-K.; Sarma, D.; Bhuyan, M.-K.; MacDorman, K.: Review of constraints on vision-based gesture recognition for human-computer interaction, *IET Computer Vision*, 12(1), 2018, 3-15. <https://doi.org/10.1049/iet-cvi.2017.0052>
- [5] Cheng, K.; Chen, Y.; Arson, K.; Rolandi, M.: Proving the value of visual design in scientific communication, *Information Design Journal*, 23(1), 2017, 80-95. <https://doi.org/10.1075/idj.23.1.09che>
- [6] Hallgreen, C.-E.; Mt-Isa, S.; Lieftucht, A.; Phillips, L.-D.; Hughes, D.; Talbot, S.: Literature review of visual representation of the results of benefit-risk assessments of medicinal products, *Pharmacoepidemiology and Drug Safety*, 25(3), 2016, 238-250. <https://doi.org/10.1002/pds.3880>
- [7] Hossan, M.-T.; Chowdhury, M.-Z.; Shahjalal, M.; Jang, Y.-M.: Human Bond Communication with Head-Mounted Displays: Scope, Challenges, Solutions, and Applications, *IEEE Communications Magazine*, 57(2), 2019, 26-32. <https://doi.org/10.1109/MCOM.2018.1800527>
- [8] Lee, H.; Kim, Y.; Bianchi, A.: A Survey on Medical Robotic Telepresence Design from the Perspective of Medical Staff, *Archives of Design Research*, 30(1), 2017, 61-71. <https://doi.org/10.15187/adr.2017.02.30.1.61>
- [9] Lou, Y.; Duan, L.-Y.; Wang, S.; Chen, Z.; Bai, Y.; Chen, C.; Gao, W.: Front-End Smart Visual Sensing and Back-End Intelligent Analysis: A Unified Infrastructure for Economizing the Visual System of City Brain, *IEEE Journal on Selected Areas in Communications*, 37(7), 2019, 1489-1503. <https://doi.org/10.1109/JSAC.2019.2916488>
- [10] Tanić, M.; Stanković, D.; Nikolić, V.; Kostić, A.: The social dimensions of space in school environment, *Facta Universitatis Series Architecture and Civil Engineering*, 16(2), 2018, 307-314. <https://doi.org/10.2298/fuace180323010t>
- [11] Wu, H.; Li, G.: Innovation and improvement of visual communication design of mobile app based on social network interaction interface design, *Multimedia Tools and Applications*, 79(1), 2020, 1-16. <https://doi.org/10.1007/s11042-019-7523-6>