Journal of Artificial Intelligence Practice (2024)
Clausius Scientific Press, Canada
DOI: 10.23977/jaip.2024.070401
ISSN 2371-8412 Vol. 7 Num. 4

Hierarchical Model of Graphical Human-computer Interface Based on Digital Twin and Visual Perception

Ruoqian Yang[1,a,*], Xingfang Peng[1,b]

[1]*School of Economics and Management, Zhongshan Polytechnic, Zhongshan, 528403, China*
[a]*yangruoqian88618@163.com*, [b]*54031525@qq.com*
[*]*Corresponding author*

Keywords: Digital Twin, Visual Perception, Human-computer Interface, Color Extraction

Abstract: With the development of science and technology in recent years, human-computer interaction has attracted more and more people's attention. Human-computer interaction not only brings great convenience to people's lives, but also can accurately recognize human images, which has been widely used in many fields. Humans are visual animals, so this paper also introduces visual perception in detail. This paper aims to study how to analyze and study the layered model of graphical human-computer interaction interface based on digital twin and visual perception. This paper proposes the basic concepts of digital twin and visual perception, and proposes a series of algorithms based on visual perception. The experimental results of this paper show that the digital twin has increased from 22% in 2015 to 47% in 2019, an increase of 25%, so it can be seen that the development of the digital twin is very rapid. Human-computer interaction has also been greatly developed, and human-computer interaction is of great significance to people's real life. Therefore, it is necessary to study the layered model of graphical human-computer interaction interface based on digital twin and visual perception.

1. Introduction

The close integration and development of new-generation information technologies like cloud computing, Internet of Things, big data, artificial intelligence, virtual reality, and traditional manufacturing has allowed countries all over the world to propose national-level manufacturing transformation strategies. Building a physical information system, realizing the connection and integration of physical factories and information-based virtual factories, and realizing intelligent manufacturing are all important goals of these methods. Digital twin has gotten a lot of attention from academic circles and businesses both at home and abroad as the greatest way to actualize the interaction and integration of physical and virtual factories.

In the design of human-computer interface, the image factor is a very important design factor. It is more pertinent to consider the evaluation and design of human-computer interface from the perspective of image. The design idea of "people-oriented" is not only reflected in the function of the product, but also in the pleasant experience generated by the user in the process of using it. Therefore, it is necessary to scientifically evaluate the image aesthetics of the interactive interface.

The innovation of this paper are: (1) This paper introduces the theoretical knowledge of digital twin and visual perception, and analyzes the importance of visual perception in the hierarchical model of graphical human-computer interaction interface using ROI extraction based on visual perception process. (2) This paper expounds the human foreground detection method of image parameters. It is found through experiments that digital twin and visual perception can improve the work efficiency of the research on the layered model of the graphical human-computer interface.

2. Related Work

Human-computer interaction has garnered academic attention as technology advances. Chakraborty B K highlighted the importance of computer vision's gesture recognition ability but did not elaborate on its limitations or provide solutions [1]. GAO reviewed vision-based gesture recognition methods, detailing their pros and cons, but did not specify the advantages of gesture recognition [2]. Michalakis K noted IoT's potential to enhance human-computer interaction but did not explain why [3]. Devi N proposed a framework for improving human-computer interaction through network-intensive technologies but lacked experimental proof [4]. Yan Jun proposed a Bayesian network model for human-computer interaction risk assessment but had not tested its feasibility in real life [5]. Li C integrated physical models and uncertainty in crack prediction using a dynamic Bayesian network, but there was no actual data to verify [6]. Zhang H proposed a digital twin-based method for rapid personalized design in manufacturing but did not explain the choice of digital twins [7]. Chen LC used deep learning for semantic image segmentation, making three contributions, but only detailed two [8].

3. Methods Based on Digital Twin and Visual Perception

3.1 Basic Concepts of Digital Twin and Visual Perception

As an effective method to realize the closed-loop control between the digital space and the physical space of the manufacturing process, digital twin is one of the effective methods to promote the intelligent development of product components and realize intelligent manufacturing. Its three parts include physical products in the real world, virtual products in the virtual world, and data and information connecting the virtual space and the real space [9]. Currently, product models are completely digitally defined in 3D, and manual methods and methods of collecting product-related information based on paper documents have been transferred to online real-time collection of RF data transmission, sensors, etc. [10]. Data transmission and laser tracking data transmission have strengthened the connection between digital space and physical space, and digital twins have gradually attracted attention. The development of digital twins from 2015 to 2019 is shown in Table 1.

Table 1: Development of digital twin from 2015 to 2019

years	Percentage%	growth rate
2015	22%	1%
2016	30%	8%
2017	36%	6%
2018	40%	4%
2019	47%	7%

As shown in Table 1, digital twins are mainly used to make digital models that are equivalent to physical entities. The virtual body can simulate and analyze the physical entity, and monitor the running status of the physical entity according to the real-time feedback information of the physical entity, and monitor the running status of the item according to the collected data. The motion data of the object improves the simulation analysis algorithm of the virtual body, and provides more accurate decision-making for subsequent operations and improvements [11].

Mechanisms of visual perception have developed into one of the most popular areas of research over the past 30 years. This scientific field was originally born out of human curiosity, but it has not been limited to theoretical exploration at the laboratory stage [12-13].

3.2 ROI Extraction Based on Visual Perception Process

For a long time, human vision, as an important process of artificial intelligence, has been concerned by researchers in related scientific fields such as machine vision technology. The biological visual system is very complex. Under the control of the visual selection strategy, it can perceive and process the ever-changing information entering the visual space [14]. The vision system can automatically compress the useless information at each stage and extract the content of interest. This gave people the inspiration to use visual computing models to simulate the visual mechanism [15].

(1) Saliency map calculation process

Inspired by the research results of lateral inhibitory neural network and receptive field in neurophysiology, the point of interest in the image is simulated by the fixation point of the human eye system. On this basis, the saliency map extraction method in the Itti computational model is used to find the fixation points.

A Gaussian pyramid is an ordering consisting of a series of related images, and images at each level will be low-filtered to produce images at the next level. The input image at the bottom of the pyramid is represented by matrix G_0, which is formula 1:

$$G_k = R(G_{k-1}) = \sum_{m=-2}^{2}\sum_{n=-2}^{2} w(m,n) G_{k-1}(2i+m, 2j+n) \tag{1}$$

Among them, w adopts the template matrix of commonly used 5×5. In order to study the frequency characteristics of the signal in the local range, some scholars put forward the "window" Fourier transform, which is the theoretical predecessor of the later famous Gabor transform. Some scholars have extended this Fourier transform function into a two-dimensional 2DGabor filter. The human eye has both multi-channel and multi-resolution visual characteristics, and Gabor filter, as a typical function with such visual characteristics, simulates human vision very well. The formula of the 2DGabor filter is as formula 2:

$$h(a,b,\theta_k,\sigma_a,\sigma_b) = \frac{1}{2\pi\sigma_a\sigma_b}\exp\left\{-\pi\left[\left(\frac{a\theta_k}{\sigma\theta_k}\right)^2\right] + \left(\frac{b\theta_k}{\sigma\theta_k}\right)^2\right\} \tag{2}$$

The above formula is a function modulated by a sine function, where σ_a is the wavelength of the sine wave and θ_k is the direction of the sine wave. The value of θ_k is formula 3:

$$\theta = \frac{\pi}{n}(k-1), k = 1,2,\dots,n \tag{3}$$

The real and imaginary parts of complex numbers have an orthogonal relationship. The real component is called a real Gabor filter, and the imaginary component is called an imaginary Gabor filter. Since the Gabor filter has both multi-scale and multi-directional attribute features, it is very suitable to use it to describe the image direction features.

From the perspective of simulating the visual cortex of the human eye to extract features, the image entering the human eye is regarded as a color channel and a direction channel, and the extraction scale is selected according to the size of the specific object in the image.

Luminance feature: use r to represent red, g to represent green, and b to represent blue, and normalize the values of the three color components so that they fall into the (0,l) interval. The luminance feature is formula 4:

$$I = \frac{r+g+b}{3} \tag{4}$$

It can be seen from the above formula that the amount of luminance information only accounts for 1/3 of the information amount of the color map. In the texture and energy calculations that do not need to consider color, the application of the luminance map saves a large amount of computation.

(2) Color feature extraction

From the point of view of the human eye observing the image, the color can directly describe the characteristics of the image. Usually, the color description of the image can be divided into two types: local and global. In general, color features are global-based features that describe the surface properties of images. Even if the direction and size of the image change, it does not affect the results of color feature extraction. The color feature extraction is shown in Figure 1.

Figure 1: Color feature extraction

As shown in Figure 1, the method of retrieval algorithm based on color features is to convert the original image to the selected color space, and then match the color space histogram according to the specific algorithm to measure the similarity.

Because the grayscale changes repeatedly in space to form the texture of an image. Therefore, if a suitable method can be found to describe the spatial relationship between pixels, it will be able to reflect the texture information of the image. This is the theoretical basis that leads scholars to study the gray level co-occurrence matrix. The gray level co-occurrence matrix is a positional relationship, from the point of gray level i to the probability of reaching $d=(Da,Db)$ at a fixed position with gray level j. The calculation formula is as formula 5:

$$p=(i,\ j,\delta,\theta)=\{(a,b)|\ f(a,b)=i,\ f(a+Da)\} \tag{5}$$

Angular second-order distance (energy): it is a measure of whether the image grayscale is evenly distributed. If the image texture is closely arranged, the angular second-order moment is larger. If the arrangement is sparse, the angular second moment f_1 is small. It is the square sum calculation of the pixel value p^2 of the co-occurrence matrix, as shown in formula 6:

$$f_1 = \sum_{i=0}^{i-1} \sum_{j=0}^{i-1} p^2(i, j) \tag{6}$$

Contrast: the contrast of an image is what makes it stand out. If the visual effect is clear, the deeper the grooves of the texture, the greater the contrast. The calculation formula is as formula 7:

$$f_2 = \sum_{n=0}^{i-1} |i-j|^2 \sum_{n=0}^{i-1} \sum_{n=0}^{i-1} p(i, j) \tag{7}$$

Wavelets have multi-scale and multi-directional characteristics, and the Gabor filter makes full use of this point feature to analyze the image. In this paper, Gabor transform is used to describe image texture features.

3.3 Human Foreground Detection Method Based on Image Parameters

This chapter proposes methods to improve the robustness of existing background subtraction methods based on image parameters. The idea of this method is: on the basis of the existing background subtraction method, the changes of image color, brightness and edge features are defined as image parameters, and the input image and foreground detection results are evaluated by the parameters of image feature changes.

Backdrop subtraction is a common moving object detection approach for detecting static or moving items that are not part of the background model. To calculate the foreground component, this approach must keep track of a background model and compare it to the current image.The structure diagram of the background subtraction method is shown in Figure 2.

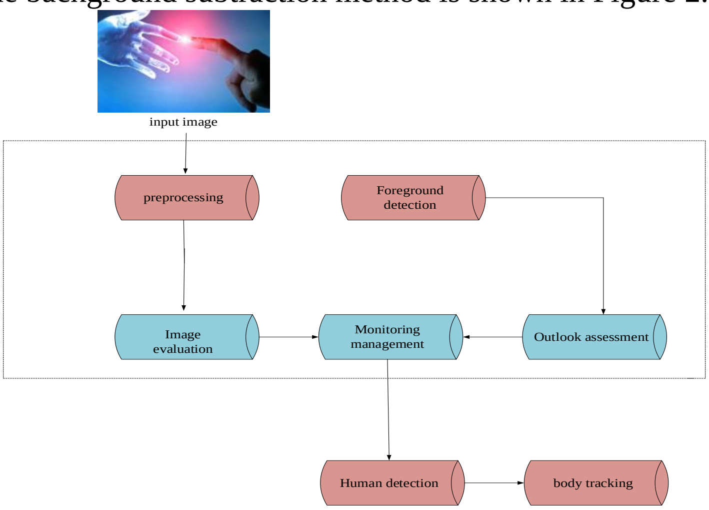


Figure 2: Background subtraction method system structure diagram

Although certain background removal methods can adapt to the dynamic background by modifying the learning rate, as illustrated in Figure 2, others cannot. However, in the interactive process, the robot requires a higher learning rate when it moves or rotates its head, and a lower learning rate while it remains immobile. As a result, finding an appropriate learning rate to adapt to all events while ensuring the accuracy of the background model is tough. To address this issue, this chapter provides a strategy for improving the resilience of existing image-based background reduction algorithms.

The color space HLS is composed of hue (H), luminance (L) and saturation (S), which is more in line with the human eye's understanding of hue, saturation and lightness.

Therefore, the image parameters in this paper are defined in the HLS color space. Image parameters include changes in image color, brightness, and edges. In the description of the following parameters, unless otherwise specified, the image refers to the image in the HLS color space.

The proportion of high-brightness pixels to all the pixels in the image. When the value is greater than the threshold $LH_{\max}$. The parameters are as in formula 8:

$$LH(t)=\frac{1}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N}LH(t,a,b) \tag{8}$$

(a,b) represents the a-th row and b-th column of the image. T_1 is a predefined threshold, when the brightness value of a pixel is greater than the threshold, the pixel is considered as a high-brightness pixel. In such scenes as turning on the lights or covering the camera with hands, the ambient brightness changes from dark to bright.

We calculate the parameter DH using the absolute value of the difference between the two image hue histograms at the same index. If a foreground pixel is moved to a new place and the hue value of its original position has changed, the absolute value of the difference is equal to 2. We calculate the sum of the absolute values of all index differences and divide by 2 to get the number of shifted pixels. Based on the above analysis, $DH(t,0)$ and $DH(t,t-1)$ are calculated by formula 9 and formula 10:

$$DH(t,0)=\frac{1}{2M\times N}\sum_{i=0}^{i}\left|h_i(t)-h_i(0)\right| \tag{9}$$

$$DH(t,t-1)=\frac{1}{2M\times N}\sum_{i=0}^{i}\left|h_i(t)-h_i(t-1)\right| \tag{10}$$

Images are often interfered by various noises in the process of acquisition and transmission. Before performing advanced image processing such as edge detection and image segmentation, it is necessary to eliminate the interference of noise.

In order to better analyze the effect of filtered image preprocessing on foreground detection, two adjacent frames of 5 groups of indoor background scenes were randomly selected for analysis. The basic conditions of two adjacent frames of five groups of indoor background scenes are shown in Table 2.

Table 2: Basic conditions of two adjacent frames of 5 groups of indoor background scenes

List	A	B	C	D	E
P_H^O	0.68	0.74	0.67	0.81	0.71
P_H^F	0.59	0.62	0.67	0.73	0.47
P_L^O	0.67	0.67	0.72	0.69	0.75
P_L^F	0.35	0.35	0.43	0.32	0.43

It can be seen from the data in Table 2 that the filtering can reduce the noise of the image, and the P_L and P_H values of the filtered image are both smaller than those of the original image. By filtering, the false detection of foreground pixels due to slight interference can be prevented. However, if the noise is very serious, filtering can only reduce the interference, but not completely eliminate the interference. In this case, the foreground is falsely detected.

3.5 Visual Receptive Field

The classical receptive field usually has a center-periphery concentric circle structure, including two parts: the excitatory area and the inhibitory area. With the in-depth study of the receptive field, a large number of electrophysiological experiments show that there is a special area around the classical receptive field that regulates the response of the central area of the classical receptive field. This regulatory effect is called the peripheral effect, and the corresponding area is called the disinhibited area of the non-classical receptive field.

According to the physiological experimental conclusions of cat retinal ganglion cells and lateral geniculate body cells, a three-Gaussian model is proposed to simulate the characteristics of the non-classical receptive field, and its mathematical expression is as shown in formula 11:

$$TG(a,b) = A\exp\left(-\frac{a^2+b^2}{\sigma_1^2}\right) \tag{11}$$

In the formula, a and b represent spatial variables; σ_1, σ_2 and σ_3 represent the size of the excitatory, inhibitory and disinhibitory regions of the non-classical receptive field, respectively. The schematic diagram of the non-classical receptive field and the response diagram of the three-Gaussian model are shown in Figure 3.

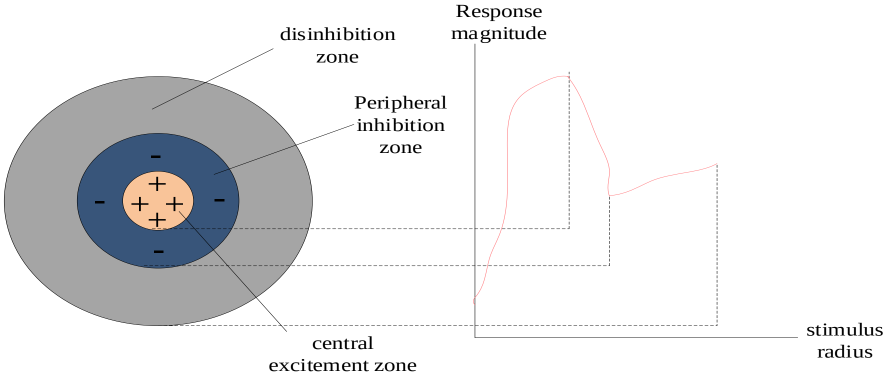


Figure 3: Schematic diagram of non-classical receptive field and response diagram of three-Gaussian model

As shown in Figure 3, the receptive fields of simple cells in the primary visual cortex had more intense responses. And it is generally believed that the receptive fields of the lateral geniculate body cells are arranged and combined according to a specific orientation, so that the receptive fields of the primary visual cortex neurons have the characteristics of direction selection.

4. Experiment and Analysis of the Layered Model of Human-computer Interaction Interface

4.1 The Evaluation of the Hierarchical Model of Human-computer Interaction Interface

With the development of society, social service robots have gradually entered people's daily life.

They provide services by interacting with users at close range in the daily environment and causing users' social behaviors. The ability to interact with humans is an important ability of this type of robot. The interactive ability is not only reflected in the robot's rich perception and response capabilities, but also requires the robot to be able to make appropriate interactive behaviors in different situations.

With the development of computer and multimedia technology, human-computer interaction has become an important research topic. The interaction based on computer, large-screen display technology, high-end electronic technology and software technology is an important direction of human research. Human-computer interaction has greatly improved the convenience for people to exchange information. The evaluation of the layered model of the human-computer interface is shown in Figure 4.

As shown in Figure 4, the evaluation of the layered model of human-computer interaction interface has the following points: human-computer interface balance, human-computer interface continuity, human-computer interface integrity, human-computer interface simplicity, etc.

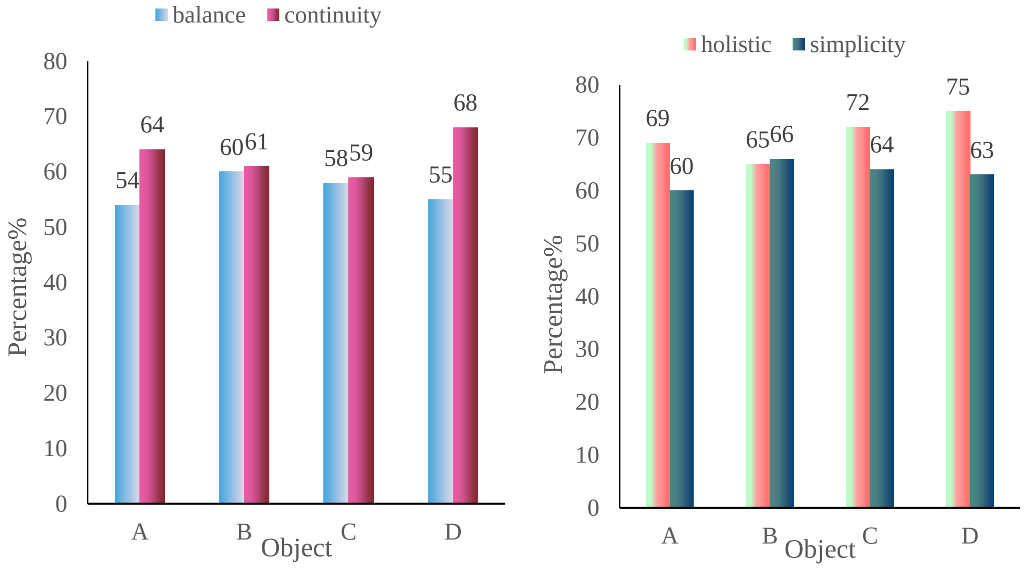


Figure 4: Evaluation of the layered model of human-computer interaction interface

4.2 Experiment and Analysis of Feature Extraction

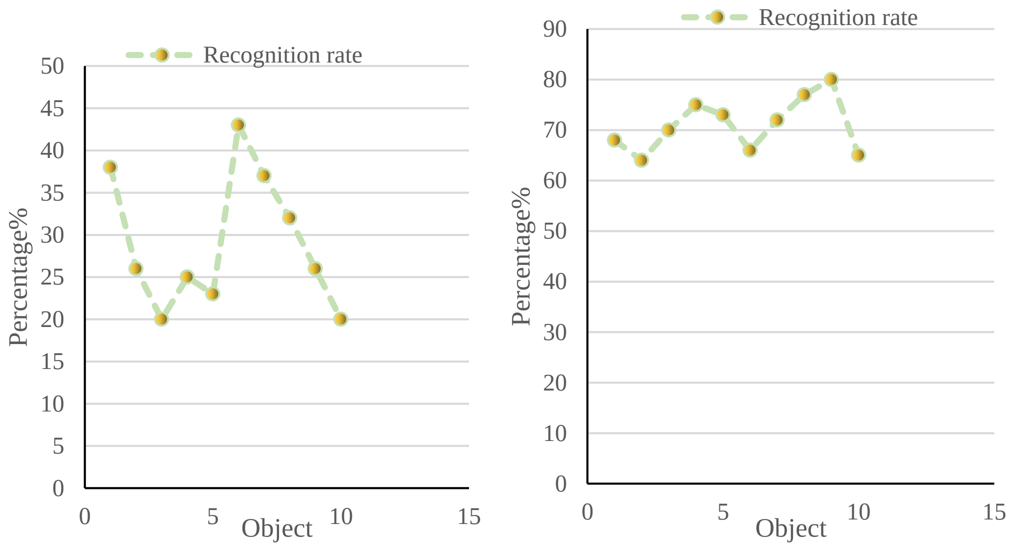
Since the number of blocks will affect the recognition rate, this paper firstly conducts experimental analysis on the number of blocks, and uses the singular value feature vector combined with the sparse representation of the face recognition method to classify and recognize it. First, it is divided into four groups: 1 block, 4 blocks, 16 blocks and 25 blocks, and the dimension of the feature vector is 64. That is, when the image is divided into 16 blocks, the pixels of each sub-block are 25*25, and the first 4 feature vectors are taken. When the image is divided into 4 blocks, the pixel of each sub-image is 50*50, and its 16 feature vectors are taken. When the image is not divided into blocks, the 25 eigenvectors corresponding to the eigenvalues of the whole image are taken. The recognition rate comparison is shown in Table 3.

As shown in Table 3, the recognition rate of the block is higher than that of the whole image, and the recognition rate increases with the increase of the number of blocks. But when it exceeds 16 blocks, the recognition rate does not improve significantly, but increases the computational complexity of recognition time.

Table 3: Recognition rates of three groups of block rates

serial number	number of blocks	Recognition rate
1	1 piece	55%
2	4 piece	61%
3	16 piece	70%
4	25 piece	84%

The comparative experiments in the face database show that the method proposed in this paper has a higher recognition rate than the traditional method. With the increase of feature dimension, the recognition rate will also increase, but the trend will become flat. The recognition rate of the common block method will be more affected with the change of the eigenvalue, while the method used in this paper has less fluctuation. Therefore, the method adopted in this paper has higher stability. The recognition rate of traditional feature extraction and the recognition rate of feature extraction based on image parameters are shown in Figure 5.



(a) Recognition rate of traditional feature extraction (Left)

(b) (b) Recognition rate of feature extraction based on image parameters (Right)

Figure 5: The recognition rate of traditional feature extraction and the recognition rate of feature extraction based on image parameters

As shown in Figure 5, when the dimension of the feature vector is low and the number of training samples is small, the recognition rate of the method proposed in this paper is more different than that of the common block method. The reason for this phenomenon is that the particle filter itself compresses and extracts the data again, so it can represent the image more effectively even when the feature dimension is small.

5. Conclusions

With the rapid development of modern science and technology, especially information technology, the information industry has become a pioneer in the world economy. With image information as the core, it provides more and more various network information services to all living and economic systems than ever before. Therefore, this paper mainly analyzes and studies the layered model of graphical human-computer interaction interface. This paper firstly introduces the general theoretical knowledge of digital twins and visual perception, and then proposes ROI extraction and human-computer interaction based on the process of visual perception based on graphics. In the experimental part, the evaluation of the layered model of the human-computer interface is firstly studied, and it is found that the human-computer interface has several dimensions of balance, continuity, integrity and simplicity. Then the experimental analysis of feature extraction is carried out, and it is found that the color extraction proposed in this paper can also identify effective images when the feature dimension is small.

Acknowledgement

This work was supported by the High-level Talent Research Initiation Project of ZhongShan Polytechnic (Research on Corporate Brand Communication Strategy from the Perspective of Digital Marketing Reconstruction, Project No. KY2210).

References

[1] Chakraborty B K, Sarma D, Bhuyan M K, et al. Review of constraints on vision-based gesture recognition for human–computer interaction[J]. Iet Computer Vision, 2018, 12(1):3-15.
[2] Gao Yongqiang, et al. Vision-Based Hand Gesture Recognition for Human-Computer Interaction A Survey[J]. Wuhan University Journal of Natural Sciences, 2020, v.25; No.130(02):82-97.
[3] Michalakis K, Aliprantis J, Caridakis G. Visualizing the Internet of Things: Naturalizing Human-Computer Interaction by Incorporating AR Features[J]. IEEE Consumer Electronics Magazine, 2018, 7(3):64-72.
[4] Devi N, Easwarakumar K S. A Clinical Evaluation of Human Computer Interaction Using Multi Modal Fusion Techniques [J]. Journal of Medical Imaging & Health Informatics, 2017, 7(8):1759-1766.
[5] Yan jun, Zhang, Youchao, et al. Safety Risk Assessment of Human-computer Interaction Behavior Based on Bayesian Network[J]. Recent Patents on Computer Science, 2017, 10(2):171-177.
[6] Li C, Mahadevan S, You L, et al. Dynamic Bayesian Network for Aircraft Wing Health Monitoring Digital Twin[J]. AIAA Journal, 2017, 55(3):1-12.
[7] Zhang H, Liu Q, Chen X, et al. A Digital Twin-Based Approach for Designing and Multi-Objective Optimization of Hollow Glass Production Line[J]. IEEE Access, 2017, 5(2017):26901-26911.
[8] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4):834-848.
[9] Kuznetsova N, Karlova T, Bekmeshov A. Building a digital twin of the main automated systems of an industrial enterprise to determine the level of information security[J]. Ergodesign, 2021, 2021(2):97-102.
[10] Granelli F, Capraro R, Lorandi M, et al. Evaluating a Digital Twin of an IoT Resource Slice: An Emulation Study Using the ELIoT Platform[J]. IEEE Networking Letters, 2021, 3(3):147-151.
[11] Zhang T, Sun S. An exploratory multi-scale framework to reservoir digital twin[J]. Advances In Geo-Energy Research, 2021, 5(3):239-251.
[12] Xu J, Guo T. Application and research on digital twin in electronic cam servo motion control system[J]. The International Journal of Advanced Manufacturing Technology, 2021, 112(1):1-14.
[13] Rakic A, Zivanovic S, Dimic Z, et al. Digital Twin Control of Multi-Axis Wood CNC Machining Center Based on LinuxCNC [J]. Bioresources, 2021, 16(1):1115-1130.
[14] Wang F Y, Li Y, Zhang W, et al. Digital Twin and Parallel Intelligence Based on Location and Transportation: A Vision for New Synergy Between the IEEE CRFID and ITSS in Cyberphysical Social Systems [Society News][J]. IEEE Intelligent Transportation Systems Magazine, 2021, 13(1):249-252.
[15] Ferguson S, Bennett E, Ivashchenko A. Digital twin tackles design challenges[J]. World Pumps, 2017, 2017(4):26-28.