Retraction

Retracted: Pose Estimation under Visual Sensing Technology and Its Application in Art Design

Journal of Sensors

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] B. Song, M. Li, and L. Lou, "Pose Estimation under Visual Sensing Technology and Its Application in Art Design," *Journal of Sensors*, vol. 2021, Article ID 2255730, 14 pages, 2021.

Research Article

Pose Estimation under Visual Sensing Technology and Its Application in Art Design

Bing Song [ID],[1] Meng Li,[2] and Liping Lou[3]

[1]*Yiwu Industrial & Commercial College, Yiwu, Zhejiang 322000, China*
[2]*Jiujiang University, Jiujiang, Jiangxi 332005, China*
[3]*Zhejiang Institute of Mechanical and Electrical Technician, Yiwu, Zhejiang 322000, China*

Correspondence should be addressed to Bing Song; songbing@ywicc.edu.cn

The study is aimed at solving the problem of large measurement errors caused by the binocular camera in traditional 3D art design, which leads to inaccurate 3D information of the target. The contour information extraction in the process of human motion pose reconstruction is easily affected by the noise in the image. Therefore, a binocular stereo vision system is built first and it integrates image acquisition, camera calibration, and image processing. The dedistortion method is used to process the image because it can reduce errors. Second, a three-dimensional human motion pose reconstruction model is implemented, the Gaussian template is used to remove the noise in the image frame, and the change detection template (CDM) is used to solve the problem of background "exposure" and "occlusion." Finally, simulation experiments are designed to verify the system and model designed. Since the research on the application of pose estimation based on visual sensing technology in art design is still blank, such research has great significance and provides a reference for the research in the field. The literature analysis is used to expound and analyze the application of pose estimation based on visual sensing technology in visual communication design and environmental art design: (1) although the binocular stereo vision system causes some errors in the measurement, the overall error is controlled within 2% and the accuracy is high, which proves that it can be applied to the acquisition of three-dimensional information of the target in art design; (2) there is a high degree of fitting between the video sequence data created by the three-dimensional human motion pose reconstruction model designed and the real motion data, which indicates that this method has high accuracy in processing image sequences and the feasibility of applying it to human pose reconstruction in three-dimensional art design is high; (3) through the analysis of the existing literature, it is found that most of the current visual-based attitude assessment studies are carried out by using network cameras combined with computers, and the quality of the obtained images is low. The combination of binocular stereo sensor and attitude estimation technology can be applied to the design of advertising, animation, games, and packaging, making the behavior of virtual characters in animation and games more vivid. The combination provides convenience for the collection of environmental spatial information and object attitude information, the formulation of a design scheme, and real-time monitoring of construction in environmental art design. The purpose of this study is to provide an important theoretical basis for the technical upgrading of art design.

1. Introduction

Art belongs to the social superstructure, and it is an important part of people's spiritual life after their needs of material life are met. If the economy is more prosperous, the higher living standards are and the greater demand for art is becoming [1]. The art design is a process in which artists express their inspiration, experience, and feelings through artworks and commu-

nicate with the public. The traditional art design process needs to go through tedious design steps and takes a lot of time, which cannot meet the practical needs of today's society. In response to the problem, the research of art design combined with science and technology attracts more and more attention of people [2]. No matter what type of art design, it needs to be conveyed through vision, such as color, pattern, and text on clothing; the quality and 3D effect of animation; the

composition of product packaging; and the shape and size of landscape [3]. With the development of society, people have higher and higher requirements for artistic products. The quality of 2D photos or video images taken by ordinary cameras is poor, and they are unable to meet practical needs. The equipment specially used for shooting ultra-high-quality films, television dramas, and animations are generally bulky, expensive, and not suitable for people to use in artistic design work.

Visual sensor based on visual sensing technology has a series of advantages, such as small size, low price, and long life. It can draw the surface signal of the object after computer processing and present it in front of the researchers. For example, the most popular binocular stereo vision sensor at present is widely used in three-dimensional modeling, three-dimensional measurement, intelligent monitoring, and other research fields [4]. Visual sensors have great advantages in image processing compared with ordinary cameras, but there are also some shortcomings, such as being vulnerable to complex background and color, occlusion, and irregular movement of the target [5]. Pose estimation refers to the estimation of the position and attitude of the target to be tested through the detection and tracking of key points. Combined with deep learning, 3D pose estimation can be realized without the interference of background and color [6]. However, traditional human motion pose reconstruction methods are susceptible to image noise when contour information is extracted. After the literature is reviewed, it is found that the current research on pose assessment based on visual sensing technology mainly focuses on human pose assessment and UAV (Unmanned Aerial Vehicle) pose estimation, but there is little literature on its application in art design.

Based on the above problems, a binocular stereo vision system and a 3D human motion pose reconstruction model are built first. Second, simulation experiments are designed to verify the system and model designed. Finally, the application of pose estimation based on visual sensing technology in visual communication design and environmental art design is analyzed. The purpose of this study is to provide an important theoretical basis for the technical upgrading of art design.

2. Materials and Methods

2.1. Analysis of Visual Sensing Technology.
Vision is the most important feeling of human beings. Through vision, the size, color, and action of objects can be perceived to obtain information about the surrounding environment. However, human visual perception is vulnerable to emotional, physical, and light, and it has certain restrictions [7]. In recent years, with the development of science and technology, visual sensors gradually replace human beings and are used in various fields, solving the problems existing in the human visual. Here, product detection is taken as an example to compare the human vision and visual sensing technology, as shown in Table 1.

The visual sensor is one of the fastest-growing and most widely used technologies in recent years. Its essence is image processing technology. It mainly uses optical devices and imaging devices to capture the image information of the external environment; that is, the image is drawn and presented in front of the researchers by intercepting the signal of the object surface [8]. The basic working principle of visual sensors is shown in Figure 1.

Visual sensing technology is mainly divided into two categories: 3D visual sensing technology and intelligent visual sensing technology. The objects seen by human eyes are hierarchical stereo images, and the shooting and display effects of 2D cannot meet the needs of people. Therefore, 3D stereo imaging becomes a research topic in recent years. Different applications such as multimedia mobile phones, robot visual navigation, automobile safety system, virtual reality, and monitoring, industrial detection are all based on 3D visual image sensor technology. The density of the reflected light on the surface of the object is different in general three-dimensional space. The three-dimensional image of the object can be obtained by detecting the density of the reflected light on the object. At present, 3D images can be obtained from solid image sensors, such as CCD (Charge Coupled Device) or CMOS (Complementary Metal Oxide Semiconductor) sensors, to detect the density of the reflected light of the object. CCD is a detecting element that transmits signals by coupling mode. It has the functions of photoelectric conversion, information storage, and transmission. CCD has such advantages as self-scanning, wide sensing spectrum range, small distortion, small volume, low system noise, low power consumption, and high reliability [9]. CMOS is a kind of chip in the computer system, and it has a low cost. Although its imaging quality is slightly lower than CCD, CMOS has the advantages of small size, low power consumption, and low cost [10]. The structure of visual image sensors is shown in Figure 2.

Chmiel et al. applied visual sensing technology to the intelligent traffic control system. A 5 million resolution CCD camera is designed for real-time analysis of the traffic situation at various intersections of urban roads and calculation of the traffic flow of each lane and queue. It can provide real-time data for the signal control system to configure the dynamic signal parameters, and realize the intelligent single-point light signal control, trunk control, and regional control. This system can be widely used in intelligent signal lamp control and traffic information collection [11].

Intelligent vision sensors under intelligent vision sensing technology, also known as intelligent cameras, are a new technology with the fastest development in the field of machine vision in recent years. The intelligent camera has the functions of image acquisition, image processing, and information transmission. It integrates image sensors, digital processors, communication modules, and other peripherals into a single camera. The integrated design makes it easy to learn, use, and maintain and has high reliability and stability, which greatly broadens the application field of visual sensing technology. The structure of the smart camera is shown in Figure 3.

2.2. Construction of the Binocular Stereo Vision System.
A complete binocular stereo vision system needs to include multiple functional modules, such as image acquisition,

TABLE 1: Comparison of human vision and visual sensing technology.

|	Human vision	Visual sensing technology
Accuracy	Need magnifying glass or microscope assistance and the accuracy is low	Accuracy can reach one-thousandth of an inch without physical constraints
Reproducibility	Easy to fatigue and has some subtle differences when testing identical products. The repeatability is poor	Using the same method and parameters to test the product. It will not fatigue and has good repeatability
Speed	Slow detection speed, especially when the product has a certain speed displacement	Faster and better detection of high-speed moving objects on production lines to improve production efficiency
Objectivity	The subjective judgment which is easily influenced by emotion when testing products	The test results are objective and reliable
Cost	Fatigue, illness, or other subjective factors may lead to lower productivity and higher costs	No illness, no need to stop, high efficiency, production efficiency, cost savings
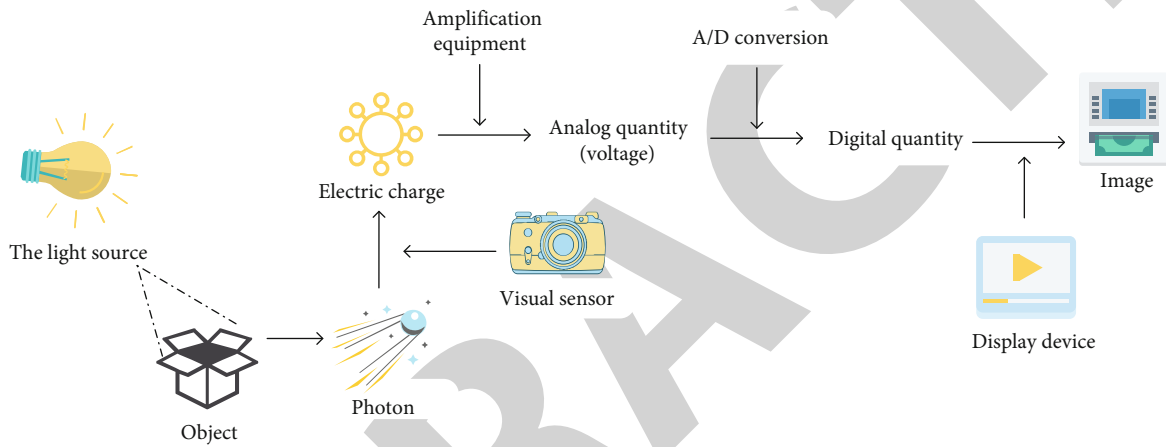
FIGURE 1: Working principle of visual sensors.
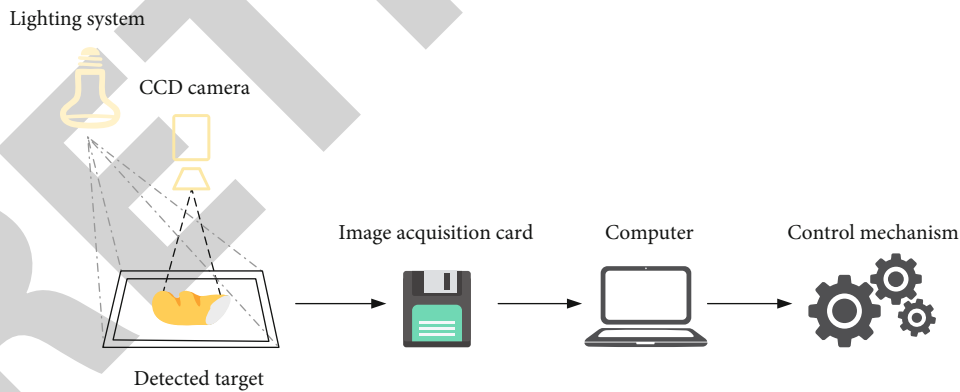
FIGURE 2: Structure of visual image sensors.

camera parameter calibration, image correction, and processing. Each module needs different algorithms to achieve its functions. Although a lot of previous studies have been carried out, there are also many new methods to improve the accuracy of the disparity map, solving the problems such as the obscured area and boundary between two given images.

2.2.1. Geometric Principle and Depth Relationship of the Binocular Vision Sensor. Binocular stereo vision technology is based on the principle of parallax and uses imaging equipment to obtain two images of the measured object from different positions. The three-dimensional geometric information of the object is obtained by calculating the position deviation between the corresponding points of the image. The principle is similar to the imaging principle of human vision. Two CCD/CMOS cameras with relatively fixed positions are used to obtain the image pairs of the measured object. The geometric model of the measured object can be established by extracting and matching the image
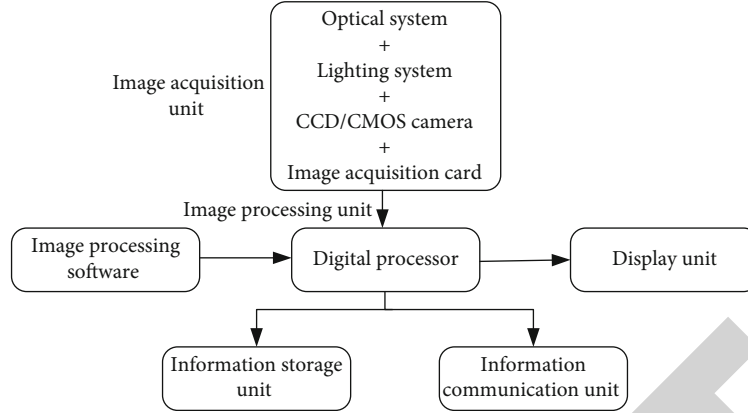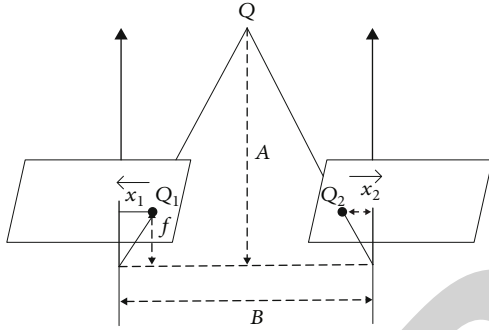
FIGURE 3: Structure of smart cameras.

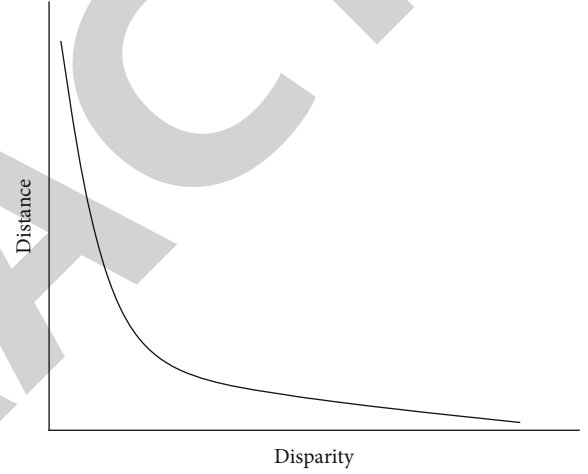FIGURE 4: Geometric principle of the binocular vision sensor.

FIGURE 5: Relationship between disparity and depth.

feature points [12]. Based on the parallax principle and similar triangle principle, the three-dimensional information of the object is obtained through multiple images. The geometric principle of the binocular vision sensor is shown in Figure 4.

In Figure 4, B is baseline distance, namely, the center distance of the left and right camera; Q_1 and Q_2 are the image point of a point Q in the space on the imaging plane of the left and right cameras; x_1 and x_2 are the distance between left and right image points and boundary of camera imaging plane; A is the distance from point Q to camera imaging plane; and f is focus on the camera.

The distance between a point in the space and the camera imaging plane can be calculated according to the proportional relationship using the principle of a similar triangle. The parallax d of point Q on the left and right cameras can be calculated by the following equation:

$$d = |x_1 - x_2|. \qquad (1)$$

The distance between Q_1 and Q_2 is set by y, and y can be obtained by the following equation:

$$y = B - (x_1 - x_2). \qquad (2)$$

From the principle of the similar triangle,

$$\frac{B - (x_1 - x_2)}{A - f} = \frac{B}{A}. \qquad (3)$$

The distance from point Q to the imaging plane of the camera can be calculated by the following equation:

$$A = \frac{fB}{x_1 - x_2}. \qquad (4)$$

Equation (4) shows that the distance from a point in space to the imaging plane of the camera is inversely proportional to the parallax of the point on the left and right cameras, as shown in Figure 5.

2.2.2. Triangulation. In general, after the calibration of the binocular camera, the image is first processed by the method of dedistortion, and the nonlinear factors causing the subsequent experimental error are excluded. After processing, the linear equations are established by using the optical triangulation method to calculate the spatial coordinates. The

imaging process of an object includes several coordinate transformations:

(1) The world coordinate system is transformed into the camera coordinate system, which can be described by a rotation matrix R and translation vector T:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = R \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + T \longrightarrow \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{5}$$

In equation (5), (x_w, y_w, z_w) is the world coordinate system and (x_c, y_c, z_c) is the camera coordinate system.

(1) The camera coordinate system is converted to the image coordinate system. This equation is obtained from the triangular proportional relationship:

$$x_u = \frac{f x_c}{z_c}, y_u = \frac{f y_c}{z_c} \tag{6}$$

In the above equation, $f x_c$ and $f y_c$ are the focus in the direction of x_c and y_c.

The matrix is expressed as

$$z_c \begin{bmatrix} x_u \\ y_u \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}. \tag{7}$$

In the equation, (x_u, y_u, z_u) is the image coordinate system.

After the first-order radial distortion is considered, the distortion model is established:

$$x_d = \left(1 + k_1 r_d^2\right) x_n, y_d = \left(1 + k_1 r_d^2\right) y_n. \tag{8}$$

In equation (8), (x_d, y_d) are the coordinates of a point on an image plane in a plane coordinate system. k_1 and k_2 are distortion parameters.

$$r_d^2 = x_d^2 + y_d^2. \tag{9}$$

The matrix is expressed as

$$\begin{bmatrix} x_d \\ y_d \\ 1 \end{bmatrix} = \left(1 + k_1 r_d^2\right) \begin{bmatrix} x_u \\ y_u \\ 1 \end{bmatrix}. \tag{10}$$

(1) The image coordinate system is converted to the pixel coordinate system
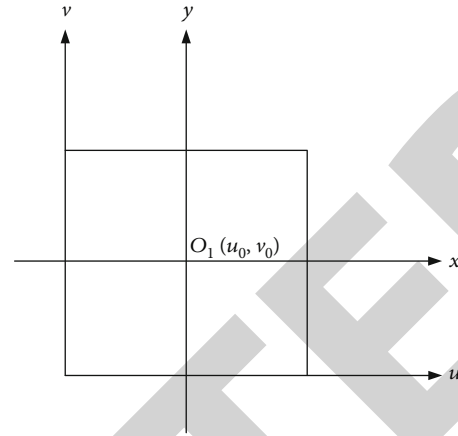

Figure 6: The pixel coordinate system.

In the coordinate system in Figure 6, the coordinate is (u, v), and $O_1(u_0, v_0)$ is the main point coordinate, so the following relationship is satisfied:

$$\begin{cases} u = \dfrac{x_d}{d_x} + u_0, \\ v = \dfrac{y_d}{d_y} + v_0. \end{cases} \tag{11}$$

In the equation, d_x and d_y are the physical sizes of pixels in the directions x and y. The pixel coordinate system is shown in Figure 6.

The matrix is expressed as

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \dfrac{1}{d_x} & 0 & u_0 \\ 0 & \dfrac{1}{d_y} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_d \\ y_d \\ 1 \end{bmatrix}. \tag{12}$$

In summary, the relationship between $x_w - y_w - z_w$ coordinate system and the $u - v$ coordinate system is obtained:

$$z \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R & T \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = K \begin{bmatrix} R & T \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}. \tag{13}$$

K is the internal parameter matrix of the camera, which needs to be obtained in the calibration:

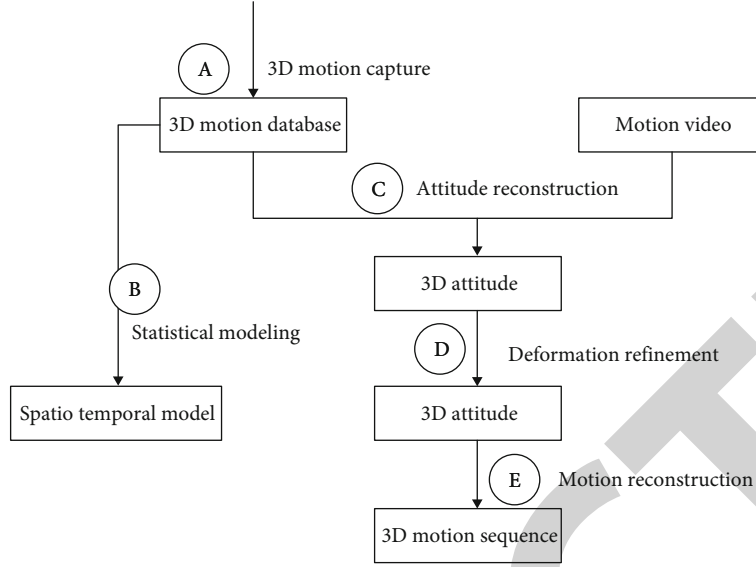$$K = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{14}$$

Figure 7: 3D human motion pose reconstruction.

In equation (14), f_x and f_y are the focal length in the directions x and y.

(1) Errors in camera calibration

Due to the radial distortion and tangential distortion of the camera, further correction is needed for the calibration results to reduce the error caused by distortion. Since the tangential distortion has little effect on the camera calibration results, the error caused by radial distortion is only considered here. Radial distortion is expressed as

$$\begin{cases} x_0 = x\left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6\right), \\ y_0 = y\left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6\right). \end{cases} \quad (15)$$

In the above equation, (x_0, y_0) are the point coordinates after distortion correction, k_1, k_2, and k_3 are radial distortion parameters, and r is the distance between (x, y) and the imaging center.

The coordinate relationship between the optical center and the imaging point is

$$\begin{cases} x_0^2 + y_0^2 = r_0, \\ x^2 + y^2 = r, \end{cases} \quad (16)$$

$$r_0 = k_3 r^7 + k_2 r^5 + k_1 r^3 + r.$$

Constructive functions can be obtained:

$$f(r) = k_3 r^7 + k_2 r^5 + k_1 r^3 + r - r_0. \quad (17)$$

2.3. Construction of 3D Human Motion Pose Reconstruction Model for Binocular Sensing Technology. 3D human motion pose reconstruction is based on the human contour information in the video to obtain the initial three-dimensional

information of human postdeformation. Then, the initial three-dimensional post is refined to obtain the accurate three-dimensional structure of human information corresponding to the video. Finally, the spatiotemporal model of the human motion database is reconstructed. The construction process of a 3D human motion pose based on the content of videos is shown in Figure 7.

(1) Image preprocessing

Due to the influence of the external environment, there are many noise points in the obtained videos and images, which will greatly reduce the accuracy of video object segmentation. Therefore, it is necessary to filter the image before segmentation to remove the influence of noise. Because the noise in the image frame is often cluttered and randomly distributed, the Gaussian template is often used to smooth the Gaussian noise. The mathematical expression of 2D Gaussian function $G(x, y)$ is shown in the following equation:

$$G(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right). \quad (18)$$

In equation (18), (x, y) represents the template coordinates of pixels. σ is the standard deviation of normal distribution.

Gaussian smoothing is conducted by Gaussian template in practice:

$$f_t(x, y) = W * I_t(x, y). \quad (19)$$

In the equation, W is the Gaussian smoothing window, $I_t(x, y)$ is the original image frame, and $f_t(x, y)$ is the smoothed image frame.

(2) Human contour extraction in the static state

Extracting foreground contour from human motion video is an indispensable step in contour-based motion analysis technology. The video does not contain the motion information of the motion camera, so in the process of processing, the static background technology is used to segment the human motion, which is easier to process. The method based on Change Detection Mask (CDM) is used to solve the video frame difference. Through the pairwise difference of the continuous three-frame video sequence images, the difference results are binarized, and then, the "and" operations are performed. The shape contour of the moving target in the middle frame can be well detected, and the problem of background "exposure" and "occlusion" can be better solved.

The three consecutive input frames are set as $f_{t-1}(x, y)$ and $f_{t+1}(x, y)$, respectively, and the absolute gray difference between the two adjacent source images is calculated:

$$
\begin{aligned}
D_{(t-1,t)}(x, y) &= |f_{t-1}(x, y) - f_t(x, y)|, \\
D_{(t,t+1)}(x, y) &= |f_{t+1}(x, y) - f_t(x, y)|.
\end{aligned}
\tag{20}
$$

Two gray difference images $D_{(t-1,t)}(x, y)$ and $D_{(t,t+1)}(x, y)$ are obtained. The difference image must be binarized first, and the appropriate threshold value is selected, which is a very important step. The approximate optimal threshold is obtained by histogram statistics:

(1) For the obtained difference image, the gray histogram is calculated to obtain the minimum gray value $I_{\min}$ and the maximum gray value $I_{\max}$, and then, the initial threshold T_0 is taken:

$$
T_0 = \frac{(I_{\min} + I_{\max})}{2}
\tag{21}
$$

(2) T_k is set as the threshold obtained after the kth iteration ($k = 0,1,2\ldots$) and used to get the gray value of the image which is initially divided into the two sets of A and B, and the average gray value of the two sets are calculated:

$$
\begin{aligned}
Z_A &= \frac{\sum_{I(x,y)<T_k} I(x, y) \times N(x, y)}{\sum_{I(x,y)<T_k} N(x, y)}, \\
Z_B &= \frac{\sum_{I(x,y)\geq T_k} I(x, y) \times N(x, y)}{\sum_{I(x,y)\geq T_k} N(x, y)}
\end{aligned}
\tag{22}
$$

In the equation, $I(x, y)$ is the gray value of pixels. $N(x, y)$ is the weight value of pixel (x, y), which is 1.

(3) Take the average of the new thresholds Z_A and Z_B:

$$
T_{k+1} = \frac{(Z_A + Z_B)}{2}
\tag{23}
$$

(4) If $T_k = T_{k+1}$, the algorithm ends and returns the obtained threshold, otherwise it goes to step 2. The threshold T obtained by the above method are used to binarize $D_{(t-1,t)}(x, y)$ and $D_{(t,t+1)}(x, y)$ to obtain two binary images $B_{(t-1,t)}(x, y)$ and $B_{(t,t+1)}(x, y)$:

$$
B(x, y) = \begin{cases} 0, & \text{if } I(x, y) < T, \\ 1, & \text{if } I(x, y) \geq T \end{cases}
\tag{24}
$$

Then, the "and" operation is performed on each corresponding pixel position to obtain the final differential binary image.

2.4. Simulation Experiment Method

2.4.1. Measurement Experiment of the Binocular Vision Sensor.
In the measurement experiment of binocular vision sensors, a scene is randomly selected. First, several points with the same distance but different positions or different positions and distances are selected as the key points of measurement. Then, the left and right images of this scene are captured by the binocular camera. After the disparity map of the image is obtained by the binocular stereo vision system, the distance of the key points in the image is measured. The measurement results are compared with the actual distance, and the error is calculated to explore the accuracy of the measurement of the binocular vision sensor system designed. The hardware components of the system designed mainly include a computer and a binocular camera. And an economically applicable binocular camera is selected, and it can restore the real color of the object well and support multiple resolutions. This camera can be applied to 3D modeling, 3D ranging, and depth detection. Therefore, the camera can meet the basic requirements of this experiment. The hardware structure of the system is shown in Figure 8.

The experimental configuration and parameter settings are as follows:

(1) Hardware configuration

In the experiment, the binocular camera is fixed on the whiteboard, and its data line is connected with the computer USB interface. The laptop used in this experiment is a Macbook pro. The processor is 2.3 GHz, four cores are Intel Core i5, and the memory is 8 GB. The specific parameters of the camera used are shown in Table 2.

(2) Software development environments

PyCharm 2019.1 professional version, Matlab 2018a, OpenCV3.6, and Python are selected as the basic software development environment for experiments. In PyCharm, the classic OpenCV library is used to process images based on Python. The NumPy module also needs to be used in image processing.

During the experiment, NumPy is repeatedly combined with SciPy and Matplotlib. In-network training, the image format of the dataset needs to be converted to a faster .npy
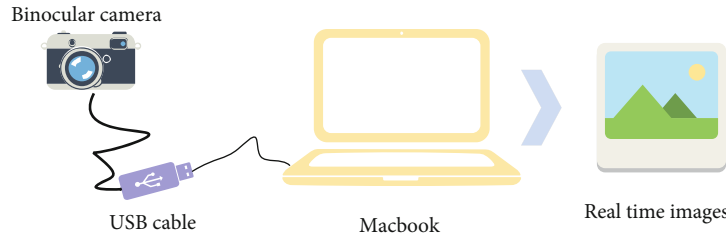
Figure 8: Hardware structure of the system.

Table 2: Specific parameters of the camera.

Camera	RER-1MP2CAM002-V90
Sensors	AE0330
Pixel dimension	$5.07 \mu m * 0.038 \mu m$
Output image format	MJPEG
Signal-to-noise ratio	39 dB
Frame rate	30 fps
Resolution	$640 * 480$ or $320 * 240$
Sensitivity	2.0 V/lux-sec@550 nm
Focal length	60 mm
Color	Color images
Aperture	Hand operate
Weight	140 g
Lowest illumination	0.1 lux

format. When the parallax of the self-shot image is predicted, the image format should also be processed first.

(3) Human motion pose reconstruction experiment

The Poser software is used to synthesize the video data sequence. First, the Gaussian template is used to remove the noise of the videos and images. Then, the static background human contour extraction method designed is used to obtain the human contour information in the video and the initial three-dimensional information of the human pose deformation The accurate three-dimensional structure of human information corresponding to the video is obtained, and the three-dimensional human motion posture reconstruction is completed. The reconstruction results are quantitatively compared with the real motion data.

2.5. Analysis of Pose Estimation Based on Visual Sensing Technology. Pose estimation refers to the problem of determining the orientation of a three-dimensional target object, which has theoretical significance and application value in tracking, control, navigation, and positioning in military and civil fields [13]. With the development of information technology, pose estimation based on visual sensor technology has gradually entered into all aspects of people's lives, and its importance has become increasingly prominent, attracting many scholars and research institutions in China and foreign countries to participate in this field [14]. In recent years, the research on pose estimation based on visual

sensing technology mainly focuses on the pose estimation of humans and UAVs, which has broad application prospects.

Human pose estimation is a basic problem in the field of computer vision research, and it is also a director of concern. The main functions are as follows: (1) action recognition: it is mainly used in human-computer interaction, intelligent monitoring, and sports medicine research; (2) animation and motion capture and augmented reality: they are mainly used in 3D movies, animation, and game production; (3) training of robots: it is mainly used in the development of intelligent robots [15].

UAV is increasingly replacing the manned system to cope with dangerous, remote, or difficult situations for manned aircraft. Accurate pose estimation is the key requirement of the UAV autonomous driving system, especially for the rotary-wing aircraft during hovering. In recent years, using pose estimation based on visual sensing technology to control UAV is a very active research field [16]. UAV usually relies on IMU (Inertial Measurement Unit) and global positioning system to provide pose and velocity information. However, the low-cost IMU has the problems of sensor bias and drift, which is vulnerable to noise interference. The high-precision IMU is usually expensive and cumbersome, which limits its application in UAV [17]. Tehrani et al. proposed a new pose estimation method based on panoramic vision sensors to reduce the drift of the aircraft. A new camera system is designed, and it consists of a CCD camera, an ultraviolet filter, and a panoramic lens. The panoramic image for pose estimation is presented in the form of UV wavelength to enhance the contrast between the sky and the ground. The effectiveness of this method is verified by comparing the visual system with IMU [18]. Visual sensors have the advantages of being lightweight, low cost, and passiveness, and they can provide information about UAV motion and the surrounding environment. Therefore, the use of visual sensors can effectively compensate for the shortcomings of inertial sensors and better obtain the pose information of UAV.

2.6. The Influence of Technological Development on Art Design. The art design is the product of comprehensive psychological activities such as human knowledge, emotion, idea, and thinking. It is the process of transmitting a certain plan, idea, and problem-solving method through art visual language. Art design involves many fields of knowledge; it not only contains the performance of aesthetics but also includes logical thinking in philosophy. It is not only the process of artists' creation but also the process of logical
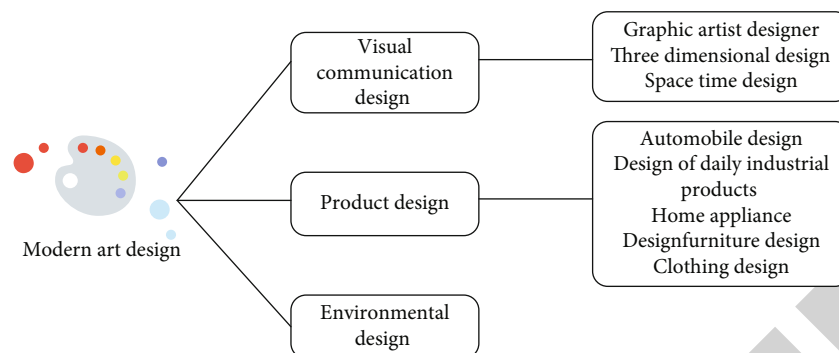
Figure 9: Classification of the art design.

thinking. Everyone has different life experiences, knowledge connotations, and different understandings of things, so the products of art design are random and perceptual and cannot be formalized [19]. From the perspective of material environment and human behavior, art design is mainly divided into three categories: (1) design for communication-visual communication design; (2) design-product design for use; and (3) design for life–environmental art design. Specific classifications are shown in Figure 9.

In the traditional way of art design, a series of complex processes such as obtaining inspiration, drawing sketches, fine creation, and coloring are needed to produce a work of art. The whole design process takes a lot of time. At present, with the development of science and technology, the application of modern science and technology in art design becomes a research topic in recent years. Liu et al. proposed a computer-based citrus peel art design to solve the problem of finding the best cutting line in citrus peel art. A designed input shape is mapped to the citrus, trying to cover the entire citrus, and the mapping boundary is used to generate the cutting path. Five customized interaction methods are developed to correct the input shape and make it suitable for citrus peel art. A large number of experiments have proved that the design and implementation method of citrus peel art based on the computer has good practicability [20]. Artificial intelligence technology can help designers get rid of tedious design steps in some ways, save design time, and improve efficiency.

3. Results

3.1. Scene Measurement Results. In the experiment, five representative points are selected from the scene disparity map for distance measurement, as shown in Figure 10.

In Figure 10, the detailed texture of the object is clear. The selected five points are located in different directions and at different distances from the lens. The measurement results are shown in Figure 11.

The above figure shows that the error is small when shooting at points 1-4 near the measurement distance, and the error is controlled within 0.6%, which has high accuracy. When point 5 is measured far away, the measurement error reaches 1.67%, which increases. The reason may be that when objects are shot far away, the deviation of measure-
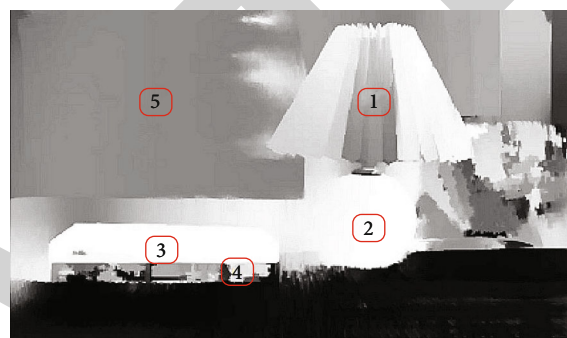

Figure 10: Scene disparity map.

ment results is largely due to the decrease of illumination conditions and clarity. However, in general, the measurement error is also controlled within 2%, which has high accuracy and is feasible.

3.2. 3D Human Motion Pose Reconstruction Based on Video Content. The quantitative comparison between reconstruction results and real motion data is shown in Figure 12.

The above figure shows that the three-dimensional human motion pose reconstruction method designed in this study has a high fitting degree between the data created by the video sequence and the real motion data. This shows that this method has high accuracy in dealing with video sequences, and the technology used in this study is based on the general processing framework, which can be used to refine various types of objects with known initial 3D information.

3.3. Application Analysis of Pose Estimation Based on Visual Sensing Technology in Visual Communication Design. Art design involves a wide range of areas, but all types of artworks need to be achieved through basic elements such as color and graphics, and this must be conveyed through vision [21]. Therefore, in art design, the vision has the advantages and status that the other four senses do not have. Visual communication design is an important part of the modern art design. It refers to the design of information communication using visual images. It is the formal rule of exploring and explaining the function, purpose, and aesthetic feeling of the art design. The three elements are text, image, and color [22]. Visual communication designs most
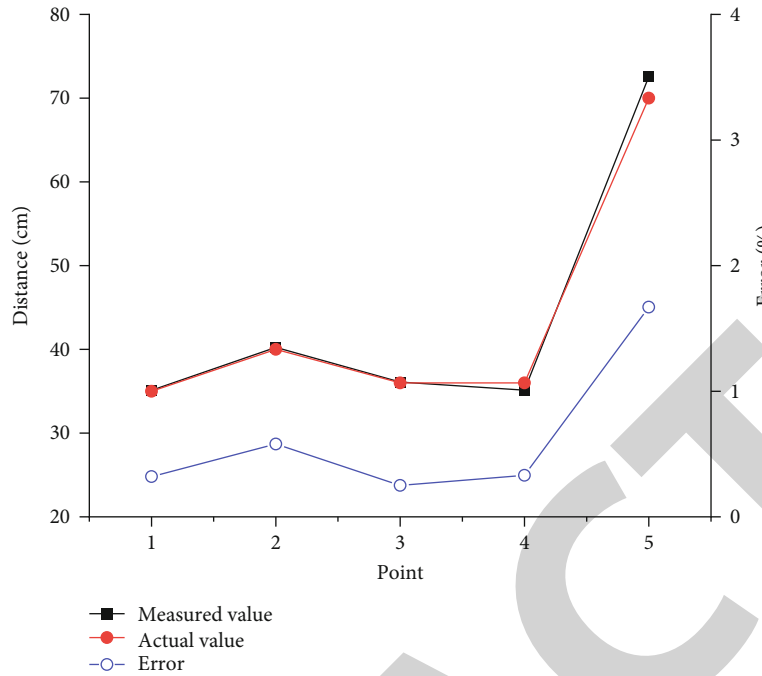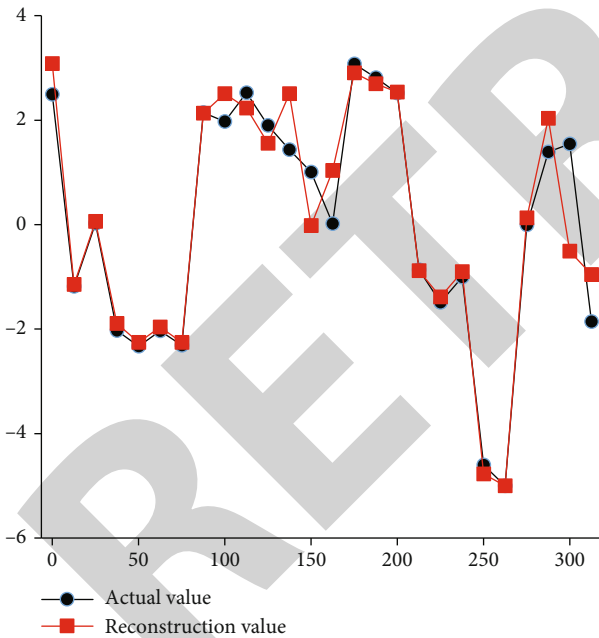
Figure 11: Scene measurement results.

Figure 12: Comparison of motion data.

close to people's lives are advertising design, packaging design, animation, and game design, and the specific classification is shown in Figure 13.

With the progress and development of the times, people's requirements for the picture and image quality of advertising, film, and television works are gradually increasing. However, the spectral range of ordinary cameras is only suitable for human vision. And after compression, the image quality is poor, which is not conducive to analysis and processing, and the dimension stays at the 2D level, which cannot meet the needs of modern society [23]. In recent years, 3D technology has been widely used in people's production and life. The market size of 3D cameras increases year by year between 2015 and 2021, as shown in Figure 14.

The 3D technology and sensors are combined recently, like binocular stereo vision sensors which are widely studied at present. Wei et al. proposed a 3D human motion capture method based on MobilePose, which is a supervised learning method for real-time detection of 2D bone joints. Due to the short time of two-dimensional joint detection and three-dimensional reconstruction, this method can realize the real-time acquisition of three-dimensional human motion. A simple human animation application program is made using the captured 3D motion information [24]. The three-dimensional space information of the target is obtained by key point detection, and the three-dimensional model is established by combining visual sensor with binocular stereo vision technology, which can be applied to the production and packaging of 3D advertising, animation, and games for better visual effects. After the two-dimensional joints are taken as the key point to estimate human posture in real time, the obtained three-dimensional position information of human joints is imported into the computer, and they should correspond to the joints of virtual characters in animation and game production. The synchronization of the two actions can make the behavior of virtual characters more vivid. Moreover, based on the characteristics of lightweight and low-cost visual sensors, it is more convenient and practical and has broad application prospects compared with professional shooting equipment.
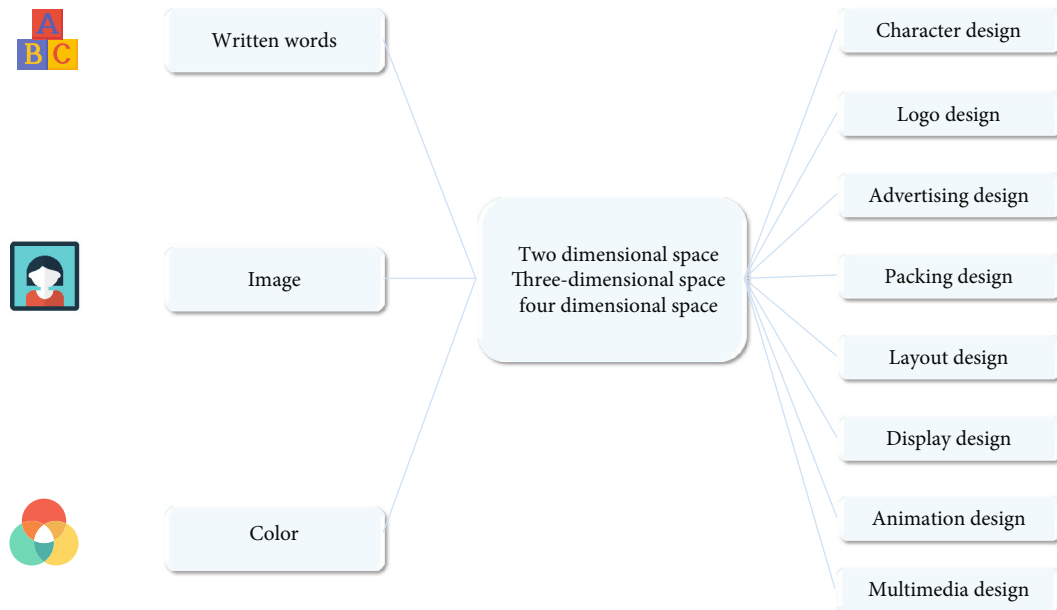
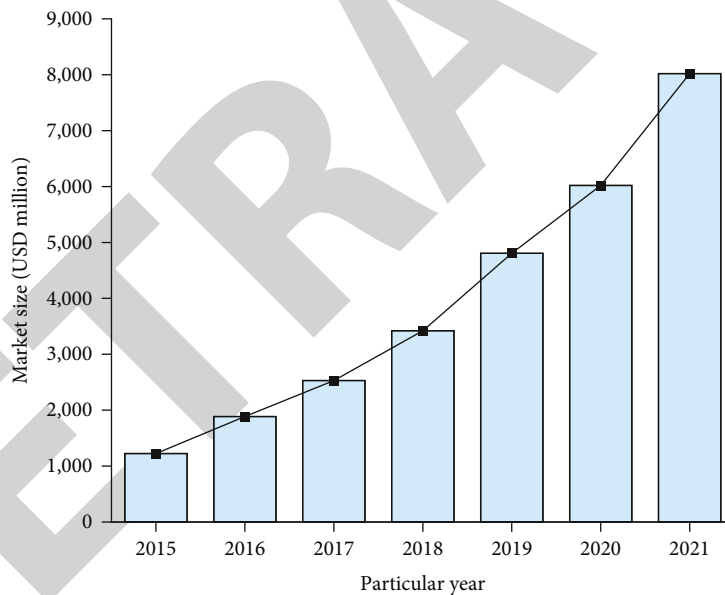FIGURE 13: Classification of visual communication design.

FIGURE 14: Market size of 3D cameras from 2015 to 2021.

3.4. The Application Analysis of Pose Estimation Based on Visual Sensing Technology in Environmental Art Design. With the improvement of people's living standards, people have higher and higher requirements for environmental quality. The concept and practice of environmental art have gradually risen and developed, becoming an extremely important scientific, cultural, and artistic achievement in China in recent years [25]. The environmental art design is to use artistic means to design human living space, coordinate the relationship between "people-building-environment," provide comfortable and beautiful space for human beings, and meet the needs of production and life in people's daily life [26]. The elements of environmental art design are shown in Figure 15.

In most of the environmental art design work, the acquisition of environmental spatial information is mainly through visual observation, manual mapping, and aerial calculation, which require a large number of human and material resources, and is also affected by factors such as weather or light. The use of double/multivisual stereo vision sensors can quickly and accurately collect environmental spatial information and object posture information and finally present the image in two or three-dimensional form, which is convenient for designers to formulate design schemes
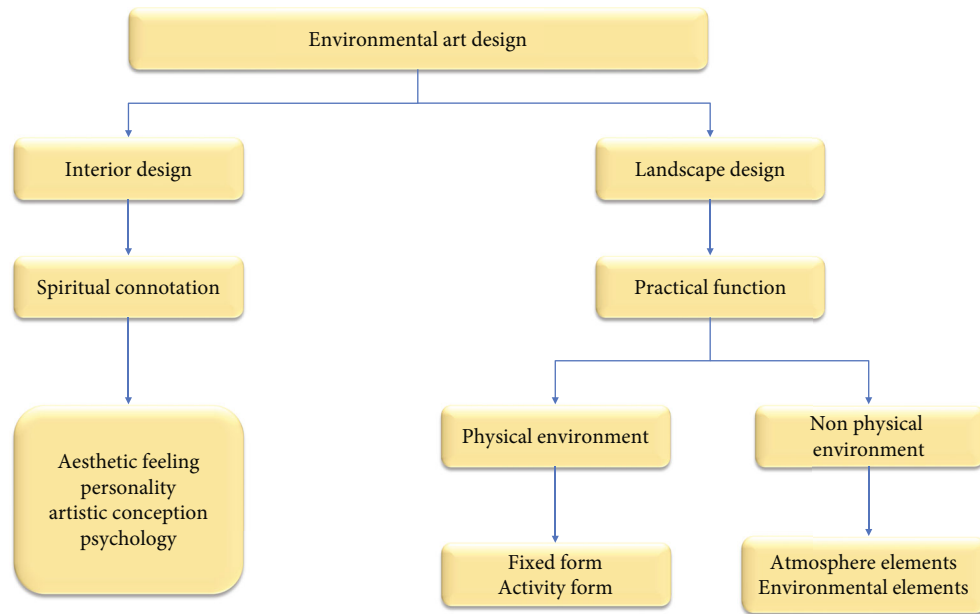
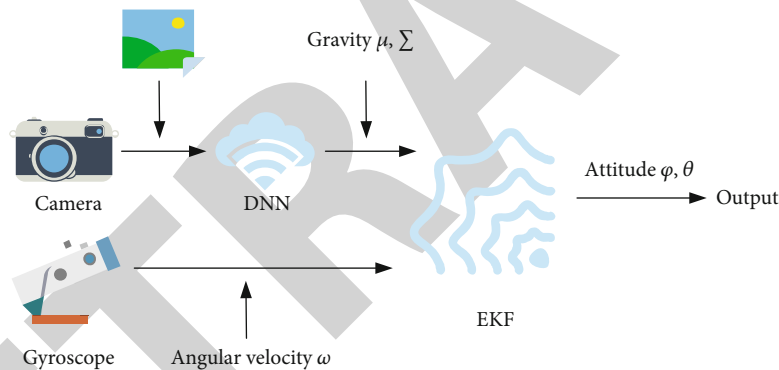FIGURE 15: Elements of environmental art design.

FIGURE 16: EKF architecture.

according to accurate information. In the process of landscape and building construction, there may be some uncertainty in the construction due to the technical differences between designers and builders [27]. For example, most artificial buildings should be built vertically, but the human eye and ordinary tools are difficult to judge some subtle differences.

Ozaki and Kuroda proposed a landscape attitude estimation method based on EKF (Extended Kalman Filter). This method uses DNN (deep neural networks) to learn terrain information and outputs the camera image in the form of the average vector and covariance matrix of gravity, which can predict the gravity vector from the single-shot image. The experimental results show that the method can predict the gravity vector from a single-shot image and obtain the attitude information of the landscape. At the same time, the use of the simulator breaks the limit of real data acquisition on the ground [28]. The EKF architecture is shown in Figure 16.

Based on the inspiration of the above method, visual sensing technology (like binocular stereo vision sensor) is combined with deep learning. The target image information is obtained by a binocular stereo vision sensor, and the three-dimensional model is established. Combined with the terrain information of deep neural network learning, the gravity vector is predicted, which can estimate the target attitude more accurately and comprehensively. This method can be applied to the monitoring of artificial landscape and building construction in environmental art design work, which ensures the safety and makes the artistic design achievements of designers perfectly presented.

In short, the combination of the binocular stereo vision sensor and attitude estimation technology can accurately obtain the three-dimensional information of the human body or the target and construct the three-dimensional model. It can also estimate the real-time target pose through the detection of the key points of the target, and it can be applied to advertising, animation, game production, and

packaging design to make it have better visual effects and make the virtual characters more vivid. Its application in environmental art design can provide technical support for the collection of environmental spatial information and object pose information, the design scheme of designers, and the real-time monitoring of scenes.

4. Conclusions

In recent years, with the rapid development of three-dimensional technology, three-dimensional art design attracts widespread attention. At present, the binocular stereo vision sensor and attitude estimation technology are widely used in this field. The former has large measurement errors and inaccurate three-dimensional information acquisition. The latter is easily affected by image noise when the contour information of the target is extracted. Based on the above problems, a binocular stereo vision system and a 3D human motion pose reconstruction model are built first. Second, simulation experiments are designed to verify the system and model designed. Finally, the application of pose estimation based on visual sensing technology in visual communication design and environmental art design is analyzed by literature analysis. The results show that (1) although there are some errors caused by the binocular stereo vision system in the measurement, the overall error is controlled within 2% and the accuracy is high, which can be applied to the acquisition of three-dimensional information of the target in art design; (2) there is a high degree of fitting between the video sequence data created by the three-dimensional human motion pose reconstruction model designed and the real motion data, which indicates that this method has high accuracy in processing video sequences and the feasibility of applying it to human pose reconstruction in three-dimensional art design is high; (3) through the analysis of the existing literature, it is found that most of the current visual-based attitude assessment studies are carried out by using network cameras combined with computers, and the quality of the obtained images is low. The combination of the binocular stereo sensor and attitude estimation technology can be applied to the design of advertising, animation, games, and packaging, which can make the design have better visual effects and the behavior of virtual characters in animation and games more vivid. It also provides technical support for the collection of environmental spatial information and objects attitude information, the formulation of design schemes, and real-time monitoring of construction in environmental art design. The deficiency of the study is that all the analysis and ideas proposed are only at the theoretical stage, and whether they are suitable for practical work remains to be further verified. The purpose of this study is to provide an important theoretical basis for the technical upgrading of art design.

Data Availability

All data analyzed during this study are included in this research article.

Conflicts of Interest

The authors declare they have no conflicts of interest.

Acknowledgments

References

[1] K. E. Maras, "Mind, language and artworks as real constraints on students' critical reasoning about meaning in art," *International Journal of Art & Design Education*, vol. 37, no. 3, pp. 530–540, 2018.

[2] Q. Wang, Z. Zhe, and Y. Xing, "Application and research of VR technology in art design teaching," *Journal of Physics: Conference Series*, vol. 1345, no. 4, article 042026, 2019.

[3] S. Li, "The trend and characteristic of AI in art design," *Journal of Physics: Conference Series*, vol. 1624, no. 5, article 052028, 2020.

[4] Y. Xu and Z. Wang, "Visual sensing technologies in robotic welding: recent research developments and future interests," *Sensors and Actuators A: Physical*, vol. 3, no. 3, article 112551, 2021.

[5] Z. Zhou, Q. M. J. Wu, F. Huang, and X. Sun, "Fast and accurate near-duplicate image elimination for visual sensor networks,," *International Journal of Distributed Sensor Networks*, vol. 13, no. 2, 2017.

[6] Á. Odry, R. Fullér, I. J. Rudas, and P. Odry, "Kalman filter for mobile-robot attitude estimation: novel optimized and adaptive solutions," *Mechanical Systems and Signal Processing*, vol. 110, no. 1, pp. 569–589, 2018.

[7] X. Ma, R. Sun, J. Cheng et al., "Fluorescence aggregation-caused quenching versus aggregation-induced emission: a visual teaching technology for undergraduate chemistry students," *Journal of Chemical Education*, vol. 93, no. 2, pp. 345–350, 2016.

[8] Z. Hou, Y. Xu, R. Xiao, and S. Chen, "A teaching-free welding method based on laser visual sensing system in robotic GMAW," *The International Journal of Advanced Manufacturing Technology*, vol. 109, no. 5-6, pp. 1755–1774, 2020.

[9] F. R. Tsai and E. Kannatey Asibu, "Modeling of conduction mode laser weldingprocess for feedback control," *Journal of Manufacturing Science and Engineering(USA)*, vol. 122, no. 3, pp. 420–428, 2000.

[10] W. Gui, "Application of sensor and detection technology in mechatronics system," *IOP Conference Series: Materials Science and Engineering*, vol. 750, no. 1, article 012111, 2020.

[11] W. Chmiel, J. Dańda, A. Dziech et al., "INSIGMA: an intelligent transportation system for urban mobility enhancement," *Multimedia Tools and Applications*, vol. 75, no. 17, pp. 10529–10560, 2016.

[12] H. Li and B. Zhang, "Application of integrated binocular stereo vision measurement and wireless sensor system in athlete displacement test," *Alexandria Engineering Journal*, vol. 60, no. 5, pp. 4325–4335, 2021.

[13] P. Narkhede, A. N. Joseph Raj, V. Kumar, V. Karar, and S. Poddar, "Least square estimation-based adaptive

complimentary filter for attitude estimation," *Transactions of the Institute of Measurement and Control*, vol. 41, no. 1, pp. 235–245, 2019.

[14] S. Wu, R. Li, Y. Shi, and Q. Liu, "Vision-based target detection and tracking system for a quadcopter," *IEEE Access*, vol. 9, pp. 62043–62054, 2021.

[15] M. A. Rizzuto, M. W. L. Sonne, N. Vignais, and P. J. Keir, "Evaluation of a virtual reality head mounted display as a tool for posture assessment in digital human modelling software," *Applied Ergonomics*, vol. 79, pp. 1–8, 2019.

[16] A. R. Vetrella, F. Causa, A. Renga, G. Fasano, D. Accardo, and M. Grassi, "Multi-UAV carrier phase differential GPS and vision-based sensing for high accuracy attitude estimation," *Journal of intelligent & robotic systems*, vol. 93, no. 1-2, pp. 245–260, 2019.

[17] L. Zhang, Z. Shi, and Y. Zhong, "Attitude estimation and control of a 3-DOF lab helicopter only based on optical flow," *Advanced Robotics*, vol. 30, no. 8, pp. 505–518, 2016.

[18] M. H. Tehrani, M. A. Garratt, and S. Anavatti, "Horizon-based attitude estimation from a panoramic vision sensor," *IFAC Proceedings Volumes*, vol. 45, no. 1, pp. 185–188, 2016.

[19] K. Budge, "Learning to be: the modelling of art and design practice in university art and design teaching," *International Journal of Art & Design Education*, vol. 35, no. 2, pp. 243–258, 2016.

[20] H. Liu, X. T. Zhang, X. M. Fu, Z. C. Dong, and L. Liu, "Computational peeling art design," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 4, pp. 1–12, 2019.

[21] G. Sachdev, "Engaging with plants in an urban environment through street art and design," *Plants, People, Planet*, vol. 1, no. 3, pp. 271–289, 2019.

[22] W. J. Liu, "The integration of contemporary art visual elements in visual communication design," *Journal of Frontiers in Art Research*, vol. 1, no. 3, pp. 4–7, 2021.

[23] R. Marroquin, J. Dubois, and C. Nicolle, "Ontology for a Panoptes building: exploiting contextual information and a smart camera network," *Semantic Web*, vol. 9, no. 6, pp. 803–828, 2018.

[24] X. Wei, Z. Zhang, S. Li, T. Yang, and X. Zhou, "Real-time 3D human motion capture from binocular stereo vision and its use on human body animation," in *Seventh International Conference on Optical and Photonic Engineering (icOPEN 2019)*, vol. 11205, Phuket, Thailand, October 2019no. 1, Article ID 112050S.

[25] A. Keller, L. Sommer, C. A. Klöckner, and D. Hanss, "Contextualizing information enhances the experience of environmental art," *Psychology of Aesthetics, Creativity, and the Arts*, vol. 14, no. 3, pp. 264–275, 2020.

[26] B. Pu, L. Bao, and K. Yang, "Research on computer 3DS MAX aided environmental art design based on performance technology and visual art," *Journal of Physics: Conference Series*, vol. 1744, no. 3, article 032040, 2021.

[27] L. da-Hong, L. Hong-Yan, L. Wei, J. J. Guo, and L. En-Zhong, "Application of flipped classroom based on the Rain Classroom in the teaching of computer-aided landscape design," *Computer Applications in Engineering Education*, vol. 28, no. 2, pp. 357–366, 2020.

[28] R. Ozaki and Y. Kuroda, "EKF-based self-attitude estimation with DNN learning landscape information," *ROBOMECH Journal*, vol. 8, no. 1, pp. 1–12, 2021.