Research Article

Research on Multicamera Photography Image Art in BERT Motion Based on Deep Learning Mode

Zhao Zhao,[1] **Mingyang Song,**[2] **and Hongyue Tang** [1]

[1]*School of Fine Arts, Hunan Normal University, Changsha 410006, China*
[2]*School of Physical Education, Hunan Normal University, Changsha 410006, China*

Correspondence should be addressed to Hongyue Tang; 202170173218@hunnu.edu.cn

In order to improve the artistic expression effect of photographic images, this article combines the deep learning model to conduct multicamera photographic image art research in BERT motion. Moreover, this article analyzes the external parameter errors caused in the calibration process and uses the checkerboard in the common field of view to calibrate the spatial coordinates of the corners of the board in multiple camera coordinate systems. In addition, this article aims to match the spatial coordinates of the corresponding points to each other and solve the rotation and translation matrix in the transformation process. Finally, this article uses the LM algorithm to optimize the calibration parameters of the camera and combines the deep learning algorithm to perform image processing. The experimental research results show that the research method of multicamera photography image art in BERT motion based on the deep learning mode proposed in this article can effectively improve the expression effect of image art.

1. Introduction

In people's daily life, as one of the main tools for image dissemination, photography has become ubiquitous to record and discover different visual possibilities. However, photography and photographic art are two completely different concepts, and photography is not related to all art. Photography is recognized as a relatively recent thing in the field of culture and art. In the 1970s, a large number of art festivals, periodicals, and galleries were launched in Western countries. Immediately afterwards, some colleges and universities set up professional photography colleges and photography departments. In addition, the research on the history of photography in the academic field is also deepening, a large number of photographic works have been included in the art collection, and more and more artists have begun to create photography. In short, the practice and dissemination of photography are no longer confined to a narrow field of practice but lead to the palace of art and culture. At the same time, people's concept of photography is constantly updated. When photography was invented, it was used only as a service tool. Today, however, it is increasingly being discussed and appreciated as the artwork itself. As a result, the public's attitude towards the practicality of photography has changed, attention has been drawn from the perceptual and rational nature of images, and the production, dissemination, and circulation of photography, as well as the form, value, and use of photography, have also changed accordingly.

Some people simply regard the art of photography as craftsmanship based on the principles of optics, chemistry, and mechanics with purity. To put it more frankly, they believe that photography is a kind of performance art of taking pictures, which leads them to ignore the lofty status of photography and its outstanding contribution to the development history of human society. In addition, the importance of photographic practice is generally ignored by some people, who only pay attention to theoretical knowledge and believe that it is enough to master the classic theories, such as visual aesthetics, cultural studies, and image expression in photographic art, and do not pay attention to photographic practice. These understandings of photography art are one-sided and not objective.

Photography art is a comprehensive artistic behavior. It is inclusive and has different connotations in different

situations. Sometimes it pays more attention to the connotation of the technical level, and sometimes it focuses more on the expression of culture and emotion. Today's photographic art is specifically a type of modern plastic art. A camera is a tool for photographic creation. Taking the photographer's creative concept as the basic starting point, they use the camera to take pictures of people or things in the real world and uses certain modern processing methods, perform artistic processing on the photographed things, and finally complete the creation of photographic works of art. Through the works showing the living conditions of human beings in the current society, at the same time, it can express the author's thoughts and feelings. Aesthetic features are more of an attribute, the presentation of their own aesthetic features. Photography, as one of many art categories, has aesthetic characteristics similar to other aesthetic activities, that is, the commonality of photography aesthetics and other aesthetic activities. In addition, photography art also has its own characteristics; it belongs to a branch of visual plastic art.

This article combines the deep learning model to study the multicamera photographic image art in BERT motion to improve the performance of photographic image art.

2. Related Work

From the perspective of the development process of image text description, it can be divided into three stages: template-based image text description method; retrieval-based image text description method; deep learning-based image text description method [1]. Before the deep learning method was proposed, most of the image description methods used template-based and retrieval-based methods. The template-based image text description method is mainly to annotate the image content, which is based on image annotation technology [2]. Template-based methods rely on visual perception of the relationship between image objects and components and describe images using representations of subject, predicate, environment, and preposition collocations [3]. Reference [4] used the method of image context subject, object, and their relationship to describe the image, used the neighbor similarity algorithm to calculate the matching degree between adjacent tuples, finally calculating the score, and the score is proportional to the matching degree. Reference [5] proposed a Conditional RaIldom Field (CRF) algorithm, the central idea of which is to generate text descriptions for predicted text labels according to template matching rules. Reference [6] improved the template in the image text description task and used the hidden Markov model to fill the template with sentences. Reference [7] applied syntactic analysis to the image text description task, used the VDR (Visual Dependency Representation) method to represent the object relationships contained in the graph with a dependency graph, and then represented the image as a VDR and then traversed the VDR, fully considering the VDR. Syntax tree relationships to fill in gaps in sentences. Template-based image text description methods may be grammatically correct in language descriptions, but the output descriptions are highly template-dependent, have

poor generalization performance, and generate text sentences that lack diversity. Due to the backwardness of this method, the image text description task is no longer used [8]. Searching for the relationship between matching text and images is the main purpose of the retrieval-based image description generation task. Retrieval-based image description also includes vision-based retrieval and multimodal-based retrieval methods [9]. The retrieval method based on visual space is to obtain textual information from the features of similar parts in the image. The image retrieval dataset established by the literature [10], for each image, uses appropriate words to describe the image. Reference [11] proposed a large dataset, which includes attribute annotations of objects, which can be used to train attribute classifiers, predicted object attributes, and improved the quality of image text descriptions. The retrieval method based on multimodal space is to perform a multimodal representation of all images and text sentences in the training corpus. For the image to be tested, retrieval is performed in the multimodal space after the image and the text are jointly mapped. First, a set of similar images of the image to be tested is obtained, and the text description corresponding to the test image is obtained according to the text description of the similar image. In [12], the authors proposed to learn multimodal space representation, use kernel function calculation method to extract high-dimensional image features, in the multimodal space represented by image and text jointly, use a sorting algorithm according to the high-dimensional features of images, and find candidate text for the set of similar images of the target. Finally, the candidate text is screened according to a certain sorting algorithm, and the text description corresponding to the image is obtained. Reference [13] applied the neural network with stronger expressive ability in the field of image text description. It makes the image text description generated in the multimodal space more accurate and of higher quality. The retrieval-based method makes full use of the dataset, but the generated image text description largely depends on the dataset. When the gap between the target image and the training data set is large, the effect of the generated text description will be poor, and only the generated text description can be generated. Human-annotated sentences are already in the dataset. Retrieval-based image description methods have better expressiveness, mobility, and practicality. Although some excellent results have been obtained, the retrieval-based method still has dependencies. The production text description is highly dependent on the training corpus, and there are problems such as high complexity, which are seriously affected by human intervention, which makes the generated text description sentences simple and ineffective. With rich semantic information, researchers continue to explore new text description methods [14]. With the popularization of deep learning knowledge, researchers have proposed new methods based on deep learning methods. The advanced and commonly used method is the end-to-end model. On the one hand, deep convolutional neural networks can be used to create models for object features in images; on the other hand, recurrent neural networks can be used to create

language models for text [15]. Reference [16] proposed a deep semantic alignment model. The model echoes images and text descriptions by aligning them. In terms of images, a region-based convolutional neural network is used for pretraining, and the features of the images are mapped to the feature space of word vectors to match the two-part features. Descriptions are generated using recurrent convolutional networks in terms of language models. The NIC model based on the English dataset proposed in [17] uses the advanced Inception V3 network to extract image features and uses LSTM to generate descriptions in the language generation model stage. This method plays an important role in the image description task. Reference [18] proposed a hard attention mechanism and a soft attention mechanism, used the combination of the attention mechanism and the LSTM network to obtain the image information of each step, and improved the expressive ability of the image description.

3. Global Calibration Optimization of Multicamera System Based on Levenberg–Marquardt Algorithm

Camera calibration has two important functions. One is to construct the relationship between the image plane coordinates and the camera space coordinates, which is explained in detail in the principle of calibration and the principle of binocular stereo imaging. The other is to construct the rotation and translation relationship between the camera coordinate system and other camera coordinate systems.

As shown in Figure 1, there are two coordinate systems in space, (o-x-y-z) and (O-X-Y-Z), and o and O are the origins of the two coordinate systems, respectively. Moreover, there is a straight line between the two origins, and the same straight line forms different vectors at different coordinates. Among them, the o point and the O point are both (0, 0.0) in their respective coordinates, and the vector from the o point to the O point is (x_1, y_1, z_1). It can also be interpreted as the coordinate of the point O in the (o-x-y-z) coordinate system is (x_1, y_1, z_1).

In the same way, the vector from point O to point o is (x_2, y_2, z_2), which is interpreted as the coordinate of point o in the (O-X-Y-Z) coordinate system is (x2, y2, 2). Since the directions of each axis corresponding to the two coordinate systems are different, these two vectors are not two opposite vectors, but the length of the same straight line is unchanged. Therefore, the magnitudes of these two vectors are the same, as shown in the following formula:

$$x_1^2 + y_1^2 + z_1^2 = x_2^2 + y_2^2 + z_2^2. \tag{1}$$

A coordinate system with two origins that do not coincide cannot discuss rotation. Therefore, we first move the (o-x-y-z) coordinate system to the vector (x_1, y_1, z_1) until its origin coincides with the origin of the (O-X-Y-Z) coordinate system.

As shown in Figure 2, in order to make the two coordinate systems completely coincide, the o-x-y-z coordinates will be rotated around the origin, and the counterclockwise
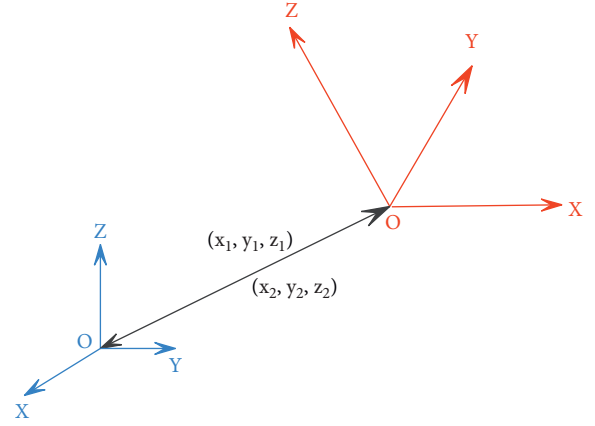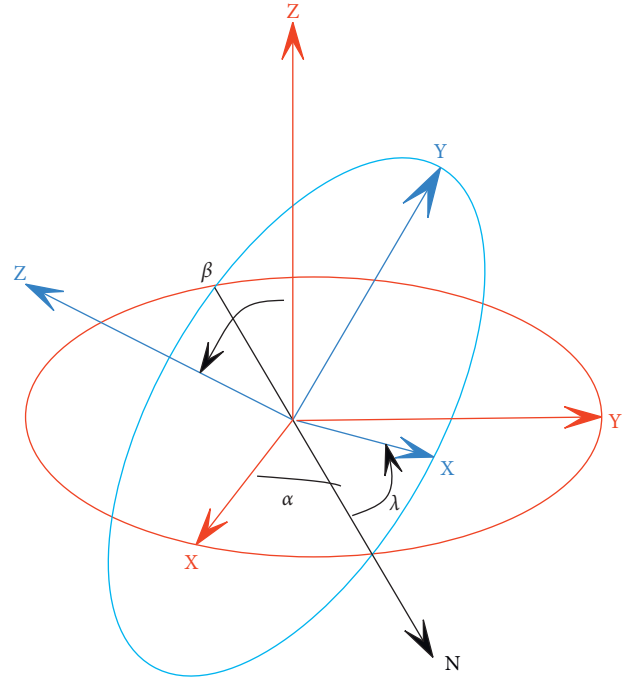

FIGURE 1: Coordinate system translation.

FIGURE 2: Coordinate system rotation.

direction is the direction of the rotation angle increment. The process is mainly divided into the following three steps.

The first step, which keeps the z-axis stationary, rotates the x-axis and the y-axis counterclockwise around the z-axis by an angle of a. At this time, x reaches the N-axis position, and a rotation variable $R_z(\alpha)$ is generated, and its matrix is shown in the following formula:

$$R_z(\alpha) = \begin{bmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{2}$$

In the second step, it keeps the x-axis stationary, which is the current N-axis, and the y-axis and the z-axis rotate counterclockwise around the x-axis by an angle of $ß$. The z-axis just coincides with the z-axis, and a rotation variable

$R_x(\alpha)$ is generated, whose matrix is shown in the following formula:

$$R_x(\alpha) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\beta & -\sin\beta \\ 0 & \sin\beta & \cos\beta \end{bmatrix}. \tag{3}$$

The third step, which again keeps the z-axis stationary, rotates the x-axis and the y-axis counterclockwise around the z-axis by a y angle. At this time, the x-axis coincides with the X-axis, the y-axis coincides with the axis, and a rotation variable $R_z(\gamma)$ is generated, whose matrix is shown in the following formula:

$$R_z(\gamma) = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4}$$

The whole process consists of three independent rotations, generating a rotation matrix R, where R is the product of $R_z(\alpha)$, $R_x(\alpha)$, and $R_z(\gamma)$, and the calculation result is shown in the following formula:

$$R = \begin{bmatrix} \cos\alpha\cos\gamma - \cos\beta\sin\alpha\sin\gamma & -\cos\beta\cos\gamma\sin\alpha - \cos\alpha\sin\gamma & \sin\alpha\sin\beta \\ \cos\gamma\sin\alpha + \cos\alpha\cos\beta\sin\gamma & \cos\alpha\cos\beta\cos\gamma - \sin\alpha\sin\gamma & -\cos\alpha\sin\beta \\ \sin\beta\sin\gamma & \cos\gamma\sin\beta & \cos\beta \end{bmatrix}. \tag{5}$$

As shown in formula (6), the two vectors (x_1, y_1, z_1) and (x_2, y_2, z_2) in the above two camera coordinate systems can be converted to each other under the action of the rotation matrix R. It can be seen that the rotation-translation matrix is invertible, and the inverse matrix of the rotation matrix is its own transpose matrix, as shown in the following formula:

$$\begin{bmatrix} -x_2 \\ -y_2 \\ -z_2 \end{bmatrix} R = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix}, \tag{6}$$

$$R^{-1} = R^T. \tag{7}$$

When multiple cameras are calibrated, due to the accuracy of target positioning and the minimization of the global error of the pursuit, as shown in formula (8), there is a very small error in the external parameters, and the same point corresponds to different coordinates in the two camera coordinate systems. Among them, the coordinate of this point in a camera coordinate is (x_m, y_m, z_m). The other camera coordinate system is converted to this camera coordinate system through the rotation and translation matrix after calibration. The corresponding coordinate is (x_n, y_n, z_n). Because it is the same point in the real space, it should coincide in the unified coordinate system. If there is no coincidence, there must be an error in the rotation and translation matrix. Therefore, this error needs to be analyzed.

If the origin of the camera coordinate system coincides after the transformation, according to the coordinate transformation process, the relative position of the points in the same coordinate system does not change, then the distance between the two points and the origin should be equal. At this time, the point coordinate can be regarded as a vector; that is, the vector's modulus is equal, as shown in the following formula:

$$x_m^2 + y_m^2 + z_m^2 = x_n^2 + y_n^2 + z_n^2. \tag{8}$$

When the origins are coincident, there is no need to retranslate the camera coordinate system, and only the parameters in the rotation matrix need to be calculated accurately. Finally, the translation vector is transformed once using (6) to reduce the error. Therefore, there is a very small angular rotation between the two coordinates, and the camera coordinate system that needs to be aligned is rotated by three angles in turn according to the rotation method shown in Figure 2.

We assume that the vectors (x_m, y_m, z_m) and (x_n, y_n, z_n) form a small angle θ. The relationship between the included angle θ and the two vectors is shown in the following formula:

$$\sin\frac{\theta}{2} = \frac{\sqrt{(x_m - x_n)^2 + (y_m - y_n)^2 + (z_m - z_n)^2}}{2\sqrt{x_m^2 + y_m^2 + z_m^2}}. \tag{9}$$

When the included angle θ is very small and approaches 0, the values of θ and $\sin\theta$ are equal, and its geometric meaning is the ratio of the distance between the two points to the vector modulus. Because the viewing angle of the camera is limited to a certain extent, in order to obtain a wider field of view, the camera will be arranged in a relatively far position, and the modulus of the camera coordinate vector will also increase accordingly. Moreover, the common field of view formed by binocular imaging is further compressed, and the measurement area will be divided into multiple segments, which will cause the measurement points in other camera coordinate systems to be converted into the global coordinate system and cause great errors. In order to avoid great errors, further optimization of the calibration external parameters must be carried out.

Another case is when the lengths of the two vectors (x_m, y_m, z_m) and (x_n, y_n, z_n) are not equal. Then, the relationship between the included angle θ and the two vectors is shown in the following formula:

$$\cos\theta = \frac{|x_m x_n + y_m y_n + z_m z_n|}{\sqrt{x_m^2 + y_m^2 + z_m^2} + \sqrt{x_n^2 + y_n^2 + z_n^2}}. \tag{10}$$

At this time, it is not only the rotation angle that causes the camera coordinate system error but also the translation

vector error. The allowable range of this error should consider the difference between the lengths of the two vectors and the distance between the two measurement points. In general, the optimization of the overall parameters of the camera during measurement will rarely cause a large change in the translation vector; that is, the distance between the origins of the two coordinate systems will not be very far. Therefore, the vector length difference will not exceed the measurement spread.

The conversion relationship between the two coordinate systems is calculated and obtained according to the geometric relationship between the coordinate systems of each camera in the process of constructing a large field of view in a multicamera system. Global calibration is to calibrate the overall measurement system to obtain the conversion relationship between all camera coordinate systems and reference coordinate systems. Taking a dual-camera station system composed of four cameras as an example, the multicamera global calibration process is described in detail with reference to Figure 3.

According to the definition of each coordinate system, it can be known that the transformation relationship between the camera coordinate systems of each camera in the entire measurement system is as follows:

The transformation of the A camera coordinate system and the B camera coordinate system of the measurement system 1 is as follows:

$$\begin{bmatrix} x_A \\ y_A \\ z_A \end{bmatrix} R_{BA} = \begin{bmatrix} x_B \\ y_B \\ z_B \end{bmatrix} + T_{BA}. \tag{11}$$

The conversion of the c camera coordinate system of the measurement system 2 to the D camera coordinate system is as follows:

$$\begin{bmatrix} x_C \\ y_C \\ z_C \end{bmatrix} R_{DC} = \begin{bmatrix} x_D \\ y_D \\ z_D \end{bmatrix} + T_{DC}. \tag{12}$$

In the above, R_{BA} and T_{BA}, R_{DC} and T_{DC} can be directly solved by the camera calibration principle. Among them, the R matrix and the T matrix can complete the inverse transformation of the two coordinate systems by formulas (6) and (7). In order to realize the coordinate conversion between the two measurement systems, the rotation matrix R_{CA} and translation vector T_{CA} of the A camera coordinate system and the C camera coordinate are directly solved by calibration. It can be solved indirectly by C camera calibration and B camera calibration rotation matrix R_{CB} and translation vector T_{CB}, or by constructing D camera and A camera calibration parameters rotation matrix R_{DA} and translation vector T_{DA}.

$$\begin{aligned} R_{CA} &= R_{BA} R_{CB}, \\ R_{CA} &= R_{DA} R_{CD}, \\ T_{CA} &= R_{DA} + T_{CD} + T_{DA}, \\ T_{CA} &= R_{BA} + T_{CB} + T_{BA}. \end{aligned} \tag{13}$$

The same target point exists in the common field of view, and the relationship between multiple target points $P_A(x_{Ai}, y_{Ai}, z_{Ai})$ in the A camera coordinate system and the corresponding target point $P_C(x_{Ci}, y_{Ci}, z_{Ci})$ in the C camera coordinate system is constructed, where $i = 1,2,3, ...$ The parameters in the transformation matrix and transformation vector that satisfy multiple target points are solved.

$$\begin{bmatrix} x_{Ai} \\ y_{Ai} \\ z_{Ai} \end{bmatrix} = \begin{bmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{bmatrix} \begin{bmatrix} x_{Ci} \\ y_{Ci} \\ z_{Ci} \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}. \tag{14}$$

If it is assumed that there is a matrix A, then A^T represents the transposed matrix of A. Similarly, $\|A\|_2$ and $\|A\|_\infty$ represent the spectral norm and row sum norm of matrix A, respectively. The singular values of matrix A are to be solved, the largest of which is the spectral norm of the matrix A. The sum of the absolute values of the elements in each row of matrix A is calculated, the largest of which is the row sum norm of matrix A.

f is a functional relationship that maps the parameter vector $P \in R^m$ to the estimated measurement vector $\widehat{x} = f(p), \widehat{x} \in R^m$. An initial estimated parameter p and a measurement vector x are provided, and it is expected to find $p +$ the vector f that best satisfies the functional relationship, that is, minimize the squared distance $\varepsilon^T \varepsilon$, where $\varepsilon = x - \widehat{x}$. The basis of the LM algorithm is to solve a linear approximation f in the neighborhood of δ_p. For a very small value of d, the Taylor series expansion approximation is as follows:

$$f(p + \delta_p) \approx f(p) + J\delta_p. \tag{15}$$

Among them, J is the Jacobian matrix $\partial f(p)/\partial p$. Like all nonlinear substitution methods, LM is iterative, starting from a chosen starting point p_0. This method generates a series of vectors $p_1, p_2, p_{3,...}$. These vectors converge to the local minima $p+$ with respect to the functional relation f. Therefore, at each step, it is necessary to find a suitable value δ_p to minimize the overall value $\|x - f(p + \delta_p)\| \approx \|x - f(p) - J\delta_p\| = \|\varepsilon - J\delta_p\|$, and finding δ_p is the solution to a linear least-squares problem. When the column space of $J\delta_p - \epsilon$ and J is orthogonal, δ_p reaches the minimum value z and $f J^T(J\delta_p - \varepsilon) = 0$ can be obtained, and the resulting value 6 can be used as the solution of the normal equation.

$$J^T J \delta_p = J^T \varepsilon. \tag{16}$$

The $J^T J$ in the matrix on the left side of the equation is an approximation to the Hessian matrix, that is, to the second derivative matrix. The LM method actually solves for small changes in the equation, which can be called augmented regular equations.

$$N\delta_p = J^T \varepsilon. \tag{17}$$

Among them, the off-diagonal elements of N are the same as the corresponding elements in $J^T J$, and its diagonal elements are $N_{ii} = \mu + [J^T J]_{ii}$, where $\mu > 0$. The process of changing the diagonal elements of $J^T J$ is called damping,
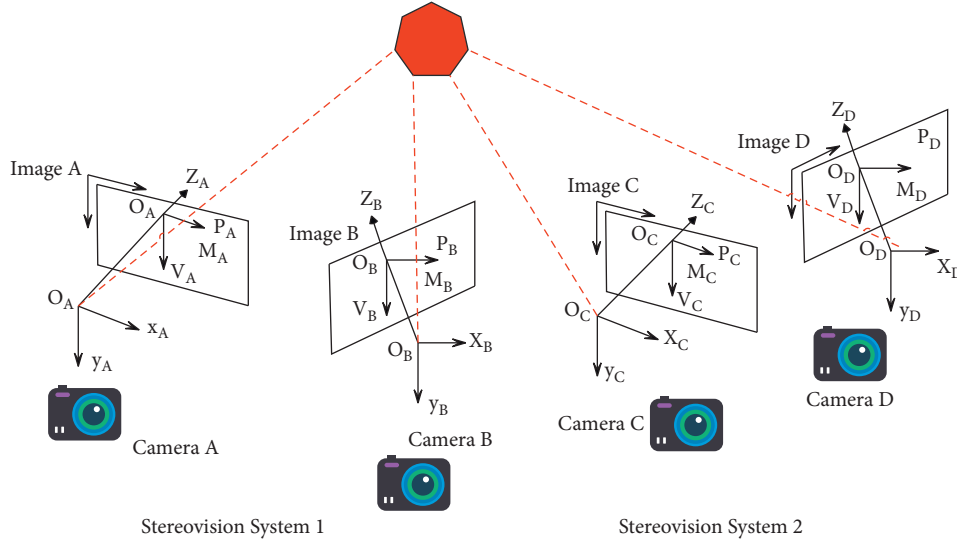
FIGURE 3: Schematic diagram of coordinate system transformation in a multicamera system.

and μ is the damping factor. In the iterative process, if a new parameter vector $p + \delta_p$ is obtained, which reduces the error ε, where δ_p can be calculated by the equation, the new parameter vector can be made closer to the optimal solution. Moreover, the iterative process is repeated with the aim of reducing the damping value.

If the new parameter vector $p + \delta_p$ is obtained and the error ε becomes larger or does not change, the damping value is increased, the augmented canonical equation is solved again, and the algorithm iterates until a value δ_p that reduces the error is found. In the equation, the solution is repeated for different damping factors μ until a parameter vector $p + \delta_p$ that is closer to the approximation is found. This update process corresponds to one iteration in the LM algorithm.

In the LM algorithm, the damping factor μ is adjusted during each iteration to ensure that the error can be reduced. If damping is set to a large value f, the matrix N in the equation used is almost on the diagonal and the LM update step value δ_p is close to the direction of the steepest descent. Also, the value of δ_p decreases in this case. Damping also handles the case of insufficient Jacobian elements, thus making $J^T J$ a singular matrix.

The LM algorithm terminates when at least one of the following conditions is met:

(1) The gradient size of $\varepsilon^T \varepsilon$ and $J\varepsilon$ on the right side of the equation will drop below a threshold.

(2) The relative change of the size of the step value δ_p falls below another threshold ε_2.

(3) The error $\varepsilon^T \varepsilon$ falls below another threshold ε_3.

(4) The maximum number of iterations $k_{\max}$ is completed.

$$J^T \sum_x^{-1} J\delta_p = J^T \sum_x^{-1} \varepsilon. \qquad (18)$$

If the covariance matrix Σ_x for the measurement vector x is available, $\sum_x^{-1}$ and the norm $\varepsilon^T \sum_x^{-1} \varepsilon$ can be incorporated into the LM algorithm by minimizing squares instead of directly solving for $\varepsilon^T \varepsilon$. Therefore, a least-squares problem with minimum weights defined by weighted canonical equations is solved.

(4) Research on multicamera photography image art in BERT motion is based on the deep learning model

The image sentiment analysis model of sample selection and image content generation based on BERT features includes four parts: "image content generation," "text feature extraction," "sample selection based on BERT features," and "image sentiment analysis." The image is processed through deep learning, as shown in Figure 4.

The specific process of image content generation is shown in Figure 5.

Figure 6 shows the flowchart of the binocular stereo vision measurement procedure. The binocular cameras are synchronously triggered to capture a frame of pictures and then transfer them to the computer memory through the USB 3.0 interface. Then, it calls the image data to perform a series of tasks such as marker point detection and image point matching with the same name, then solves the pixel coordinates of the center of the marker point, and then performs 3D reconstruction to obtain the coordinates of the marker point in physical space and output the data.

According to the actual measurement needs, a CCD camera bracket and a fixed platform, lighting equipment, etc. are added to the system as auxiliary. The connection diagram of the main part of the system is shown in Figure 7.

The camera coordinate system and the imaging plane coordinate system are established. As shown in Figure 8, the world coordinate system in this section is selected on the target plane containing low-rank textures. For the convenience of description, the following two definitions are made first. When there is no rotation of the camera imaging plane coordinate system relative to the xw0wYw plane of the world coordinate system, the image captured at this time is called a
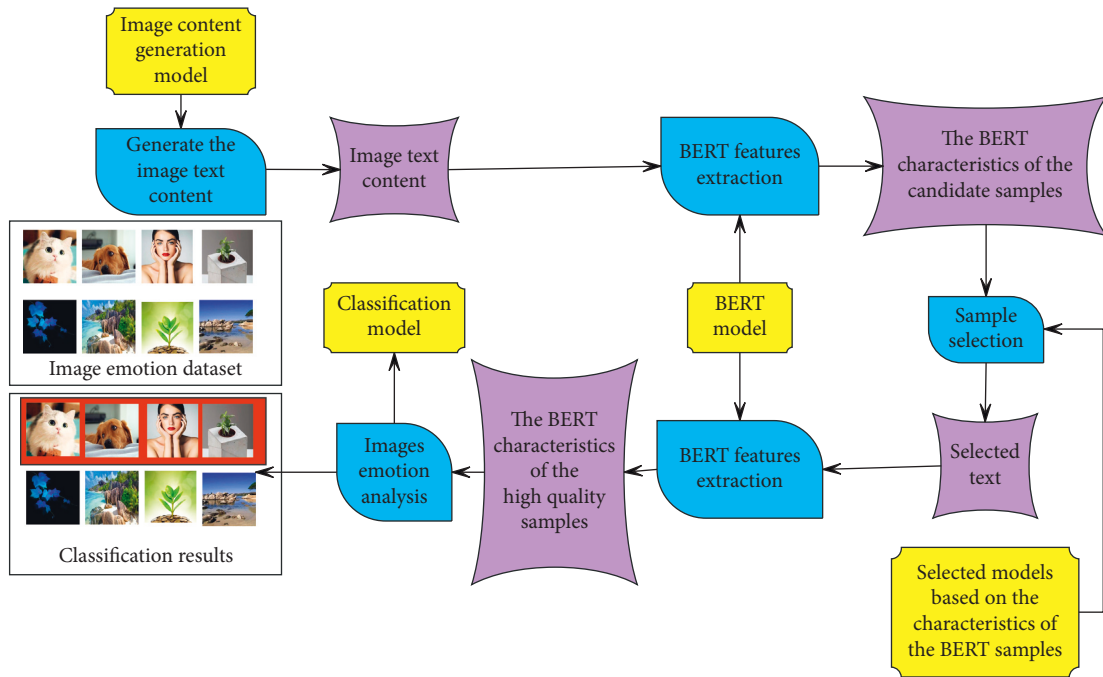
Figure 4: Image sentiment analysis process based on BERT feature sample selection and image content generation.
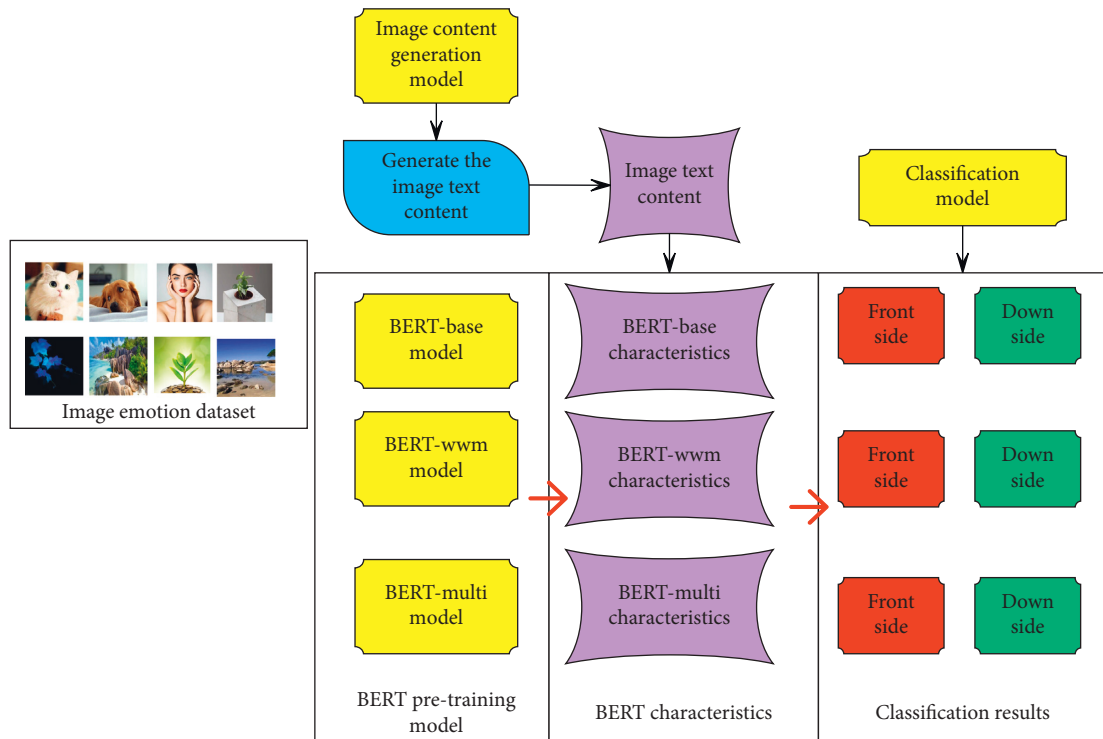

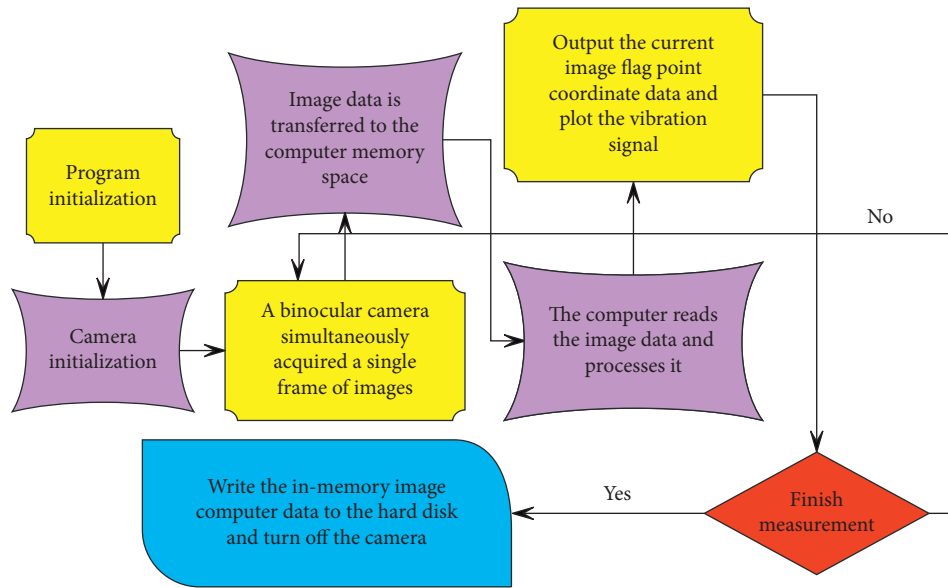Figure 5: Flowchart of the use of the BERT model.

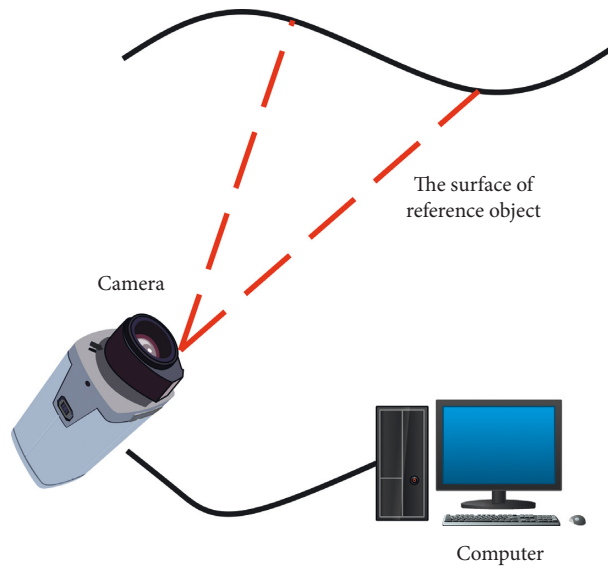FIGURE 6: Program flowchart.

FIGURE 7: Schematic diagram of camera pose measurement system.
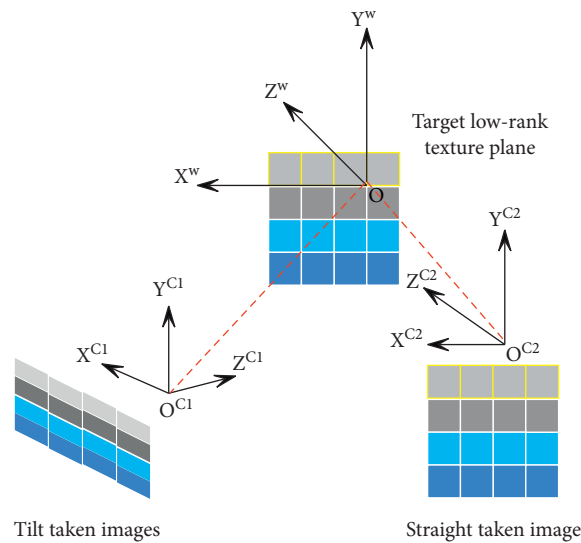

FIGURE 8: Front-facing images shot and oblique shot images.

TABLE 1: Expression effects of images obtained by multicamera photography.

Num	Image expression
1	88.18
2	88.99
3	89.51
4	92.73
5	93.77
6	93.81
7	90.25
8	93.21
9	95.29
10	94.59
11	89.42
12	90.74
13	94.84
14	91.28
15	93.75
16	89.30
17	90.44
18	90.00
19	88.68
20	95.94
21	93.01
22	95.17
23	95.38
24	89.76
25	92.51
26	94.99
27	94.53
28	90.61
29	95.25
30	91.77
31	90.05
32	89.88
33	90.56
34	89.46
35	89.66
36	90.74
37	91.43
38	92.86
39	95.43
40	89.85
41	94.19
42	95.16
43	94.74
44	88.54
45	91.94
46	95.36
47	88.43
48	93.64
49	95.12
50	89.44
51	95.78

TABLE 2: Artistic effects of images obtained by multicamera photography.

Num	Graphic arts
1	84.07
2	83.00
3	90.76
4	90.02
5	92.86
6	83.46
7	83.77
8	91.17
9	92.85
10	88.45
11	85.95
12	83.05
13	82.56
14	89.53
15	91.53
16	88.21
17	91.44
18	92.89
19	85.75
20	91.94
21	91.70
22	88.77
23	82.79
24	86.10
25	86.03
26	87.87
27	87.52
28	85.69
29	88.45
30	86.89
31	92.98
32	85.11
33	83.83
34	88.94
35	87.49
36	84.71
37	83.35
38	92.65
39	89.39
40	92.69
41	90.92
42	90.18
43	87.85
44	82.91
45	90.89
46	88.62
47	84.07
48	84.70
49	85.80
50	91.05
51	90.76

facing image. When the camera imaging plane coordinate system and XwDwYw, the plane has some unknown rotation and translation, the image captured at this time is called an oblique captured image. Obviously, compared with the original low-rank texture image, the front-facing image has no deformation, only scaling, and still retains the low-rank characteristic. When shooting at an angle, the image no longer retains low-rank properties due to projection distortion.

On the basis of the above research, the system model proposed in this article is verified, and the expression effect and image art of the multicamera image are evaluated, and the results shown in Tables 1 and 2 are obtained.

From the above research, it can be seen that the research method of multicamera photography image art in BERT motion based on the deep learning model proposed in this article can effectively improve the expression effect of image art.

4. Conclusion

Photography is a kind of visual art, but its expression must be carried out through forms, and different forms bring completely different visual experience. There are many forms of expression, and as one of the visual arts effects, ordering cannot be underestimated. "Ordering" as a guideline in graphic design can make the design more organized and normative. Similarly, for photography, "ordering" will make the photographic work have different formal meanings and show a strong sense of design order. There are many aspects to the formal expression of order in photography, such as symmetry and balance, repetition, and gradual change. These forms of order are often used in graphic art design, which can make the designed picture produce a visual experience of different orders. This article combines the deep learning model to conduct multicamera photographic image art research in BERT motion. The experimental research results show that the research method of multicamera photography image art in BERT motion based on the deep learning mode proposed in this article can effectively improve the expression effect of image art.

Data Availability

The labeled datasets used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] A. Martínez-González, M. Villamizar, O. Canévet, and J. M. Odobez, "Efficient convolutional neural networks for depth-based multi-person pose estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 11, pp. 4207–4221, 2019.

[2] M. Li, Z. Zhou, and X. Liu, "Multi-person pose estimation using bounding box constraint and LSTM," *IEEE Transactions on Multimedia*, vol. 21, no. 10, pp. 2653–2663, 2019.

[3] J. Xu, K. Tasaka, and M. Yamaguchi, "[Invited paper] fast and accurate whole-body pose estimation in the wild and its applications," *ITE Transactions on Media Technology and Applications*, vol. 9, no. 1, pp. 63–70, 2021.

[4] G. Szűcs and B. Tamás, "Body part extraction and pose estimation method in rowing videos," *Journal of Computing and Information Technology*, vol. 26, no. 1, pp. 29–43, 2018.

[5] R. Gu, G. Wang, Z. Jiang, and J. N. Hwang, "Multi-person hierarchical 3d pose estimation in natural videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 11, pp. 4245–4257, 2019.

[6] M. Nasr, H. Ayman, N. Ebrahim, R. Osama, N. Mosaad, and A. Mounir, "Realtime multi-person 2D pose estimation," *International Journal of Advanced Networking and Applications*, vol. 11, no. 6, pp. 4501–4508, 2020.

[7] N. T. Thành and P. T. Công, "An evaluation of pose estimation in video of traditional martial arts presentation," *Journal of Research and Development on Information and Communication Technology*, vol. 2019, no. 2, pp. 114–126, 2019.

[8] I. Petrov, V. Shakhuro, and A. Konushin, "Deep probabilistic human pose estimation," *IET Computer Vision*, vol. 12, no. 5, pp. 578–585, 2018.

[9] G. Hua, L. Li, and S. Liu, "Multipath affinage stacked—hourglass networks for human pose estimation," *Frontiers of Computer Science*, vol. 14, no. 4, pp. 1–12, 2020.

[10] K. Aso, D. H. Hwang, and H. Koike, "Portable 3D human pose estimation for human-human interaction using a chest-mounted fisheye camera," in *Proceedings of the Augmented Humans Conference 2021*, pp. 116–120, Rovaniemi, Finland, February 2021.

[11] D. Mehta, S. Sridhar, O. Sotnychenko et al., "VNect," *ACM Transactions on Graphics*, vol. 36, no. 4, pp. 1–14, 2017.

[12] S. Liu, Y. Li, and G. Hua, "Human pose estimation in video via structured space learning and halfway temporal evaluation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 7, pp. 2029–2038, 2018.

[13] S. Ershadi-Nasab, E. Noury, S. Kasaei, and E. Sanaei, "Multiple human 3d pose estimation from multiview images," *Multimedia Tools and Applications*, vol. 77, no. 12, pp. 15573–15601, 2018.

[14] X. Nie, J. Feng, J. Xing, S. Xiao, and S. Yan, "Hierarchical contextual refinement networks for human pose estimation," *IEEE Transactions on Image Processing*, vol. 28, no. 2, pp. 924–936, 2018.

[15] Y. Nie, J. Lee, S. Yoon, and D. S. Park, "A multi-stage convolution machine with scaling and dilation for human pose estimation," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 13, no. 6, pp. 3182–3198, 2019.

[16] A. Zarkeshev and C. Csiszár, "Rescue method based on V2X communication and human pose estimation," *Periodica Polytechnica: Civil Engineering*, vol. 63, no. 4, pp. 1139–1146, 2019.

[17] W. McNally, A. Wong, and J. McPhee, "Action recognition using deep convolutional neural networks and compressed spatio-temporal pose encodings," *Journal of Computational Vision and Imaging Systems*, vol. 4, no. 1, p. 3, 2018.

[18] R. G. Díaz, F. Laamarti, and A. El Saddik, "DTCoach: your digital twin coach on the edge during COVID-19 and beyond," *IEEE Instrumentation and Measurement Magazine*, vol. 24, no. 6, pp. 22–28, 2021.