Mobile Terminal Video Image Fuzzy Feature Extraction Simulation Based on SURF Virtual Reality Technology

YINGYING FENG

College of Information Engineering, Fuyang Normal University, Fuyang 236041, China

e-mail: fengyingying1982@yeah.net

ABSTRACT Extracting the fuzzy feature of the mobile video image can effectively improve the low illumination image quality. Traditional methods are used to construct fuzzy feature indexes of mobile terminal video images, and the detailed information of video images is divided, but the bidirectional matching of feature points is ignored, which leads to low extraction accuracy. Therefore, this paper proposes a method for extracting fuzzy features of mobile terminal video images based on SURF-based virtual reality technology. First, perform video image grayscale extraction on the input mobile terminal video image, and detect the closed area in the mobile terminal video image as the radiation invariant area of the terminal video image. Secondly, Hessian matrix is used to detect the feature points of the image, and the non-maximum suppression method and interpolation operation are used to find and locate the extreme value points. Then, the main direction of feature points was determined, and SURF description operator was used for matching to obtain initial matching point pairs. Finally, the obtained fuzzy feature one-way matching result of the video image is matched in two directions, the closest distance ratio is used to match the feature points, and the full constraint condition is used to filter out the wrong matching point pairs, thereby completing the mobile terminal video image fuzzy feature extraction. The experimental results show that the proposed algorithm is effective in feature extraction and matching, stability and speed. The misrecognition rate of the algorithm in this paper is 0.101, and the time used is only 0.41 s, which fully meets the real-time requirements.

INDEX TERMS Feature extraction, mobile terminal, video image, SURF, virtual reality technology.

I. INTRODUCTION

Virtual reality technology can combine processed information with images in real life, giving people an illusion of being in a simulated environment and meeting people's requirements [1], [2]. As a synthesis of various technologies, virtual reality is widely used in medical science, entertainment, military aerospace, interior design and other fields [3], [4]. Computer video image processing technology lays a solid foundation for the complex analysis, decision-making, and management of mobile devices with intelligent operations to extract reality technology [5]. Since the mobile device has designed a strong video image processing system in

the development stage, it has embedded operating structure, touch display, GPS positioning and video recording functions, which provides the basic conditions for the development of the extraction of the reality system to the mobile terminal [6], [7]. In the process of using current video image extraction methods to extract video images of mobile terminals, pixels with low probability may be excessively merged due to insufficient lighting, which easily leads to a reduction in the gray level of the video image and loss of detailed information of the video image of the mobile terminal. In this case, how to effectively improve the low-illumination of image quality has become a major problem to be solved urgently in the field of mobile device video image processing research and development, and has received widespread attention.

The associate editor coordinating the review of this manuscript and approving it for publication was Zhihan Lv .

Literatures [8]–[10] proposed the use of wavelet transform to achieve fast image matching, which has good time-frequency localization, multi-resolution analysis and suppression of high-frequency noise advantages, but the image feature extraction is subject to the selected wavelet base greater impact. Tang *et al.* [11] proposed Harris corner detection feature matching algorithm, which is a kind of corner extraction directly based on gray image, with high stability and sensitivity to l-shaped corner detection. But the calculation speed is slow, there are corner information loss, position shift, and clustering phenomenon. On this basis, Seada *et al.* [12], [13] improved the Harris matching algorithm, but it cannot adapt to the problem of image scale changes. Wang *et al.* [14] proposed SIFT algorithm. The method finds the extremum point in the scale space and extracts the invariant features of position, scale, and rotation. However, because the SIFT algorithm does not consider the geometric constraint information of the space, the mismatch rate is high, and obvious mismatches and mismatches are prone to appear. Wang *et al.* [15] proposed a SURF extraction algorithm. The method is improved from the SIFT feature. Through the combination of Harris feature and integral image, the running speed of the program is greatly accelerated. It can maintain invariance under both image scale and affine transformation, better robustness, but also has the problem of high false match rate. He *et al.* [16], [17] combined with other algorithms to further improve SURF, and achieved good results. Li *et al.* [18] proposed a fuzzy feature extraction method of mobile terminal video image based on video image matching. Hashemzadeh *et al.* [19] proposed a method for extracting fuzzy features of video images of mobile terminals in the form of video image denoising. This method first divides the detailed information of the video image into smooth areas, edges, corners, and isolated noise points by constructing the fuzzy feature index and gradient variation index of the mobile terminal video image. The fuzzy feature value of the diffusion tensor is set according to the result of the detailed information division. This method has a good denoising effect on video images, but there is a problem that the extraction process takes too long. Al-wswasi *et al.* [20] proposed a method for extracting fuzzy features of mobile terminal video images based on video image fuzzy feature block. This method constructs a mobile terminal video image fractional differential mask operator, defines the fractional order according to the result of the mobile terminal video image fuzzy feature block, and forms a fractional order matrix. On this basis, the fractional order the matrix is substituted into the mask operator to complete the extraction of fuzzy features of the video image of the mobile terminal. This method only extracts the details of the terminal video image to a greater extent, and the extraction effect is not ideal for the problem of holes in the video image of the mobile terminal.

A method based on SURF-based virtual reality technology for mobile terminal video image fuzzy feature extraction is proposed. First, perform video image grayscale extraction on the input mobile terminal video image, and detect the closed area in the mobile terminal video image as the radiation invariant area of the terminal video image. Secondly, the Hessian matrix and the SURF description operator are used for matching to obtain the initial matching point pair. Finally, the obtained fuzzy feature one-way matching result of the video image is matched in two directions, thus the fuzzy feature extraction of mobile video image is completed.

II. THE PRINCIPLE OF VIDEO IMAGE EXTRACTION FOR MOBILE TERMINALS

A. VIDEO IMAGE FEATURE EXTRACTION

Image feature recognition refers to the use of computers to recognize iconic attribute information in an image. These attributes include natural features and man-made features [21]. Natural features refer to direct image features such as the color, brightness, and texture of the image, and artificial features refer to image features that are indirectly obtained through calculations such as the histogram and frequency spectrum of the image. Feature extraction is to analyze and process the iconic attribute information of the image, and extract the relatively stable information that has nothing to do with random factors, such as illumination, angle, field of view and other factors, as image features [22].

In the process of extracting fuzzy features of mobile terminal video images under virtual reality technology, a detection algorithm is used to detect local fuzzy features of mobile terminal video images. After obtaining the video image fuzzy feature points, the mobile terminal video image fuzzy feature description is performed. The conventional maximum suppression method is selected to eliminate the non-blurring feature points of the mobile terminal video image. Finally, the fuzzy feature vectors of these mobile terminals are analyzed for classification and matching. Select the corresponding parameters for the fuzzy feature extraction processing of the mobile terminal video image to achieve the purpose of adaptive matching of the extraction algorithm. The specific process is as follows:

1) Assuming that the gray value of the mobile terminal video image pixel of the point to be detected is quite different from the gray value of the candidate point. In the center coordinates of a certain point in the two-dimensional mobile terminal video image, use the following formula to describe the fuzzy feature point of the mobile terminal video image:

$$Feature = \begin{cases} dark & I_{gray} \le I_{center} - threshold \\ similar & I_{center} - threshold \le I_{gray} \\ & \le I_{center} + threshold \\ bright & I_{center} - threshold \le I_{gray} \end{cases} \quad (1)$$

In the formula, t represents the threshold, which defines the gray level at any point of the video image of the mobile terminal as I_{gray} and the gray level at the center of the circle as I_{center}. When the gray value I_{center} of the central pixel is smaller than the gray value $I_x + threshold$ of the pixel at the surrounding circle point x, the gray pixels of the video

image of the mobile terminal are considered to be darker, and $similar_{I_{center} \rightarrow I_{gray}} = d$ is satisfied. Use this as a basis to derive s and the brighter grayscale pixel b.

2) Obtain candidate fuzzy feature points by comparing the gray values of pixels on the video image radius of the fixed mobile terminal. Assume that I_{center} represents the fuzzy feature point as true, otherwise false. Therefore, whether a pixel of a mobile terminal video image is a fuzzy feature point is determined.

3) Finally, train the fuzzy feature points according to the decision tree, calculate the corner response function F_V, and define a response function using the following formula:

$$F_v = \max(|I_{I_{center} \rightarrow I_{gray}} - I_{gray}| - threshold,$$
$$I_{gray} - I_{I_{center} \rightarrow I_{gray}}| - threshold) \quad (2)$$

Usually, the conventional maximum value suppression method is used to eliminate the non-blurring feature points of the mobile terminal video image. Figure 1 is a framework based on feature extraction of mobile terminal video images.
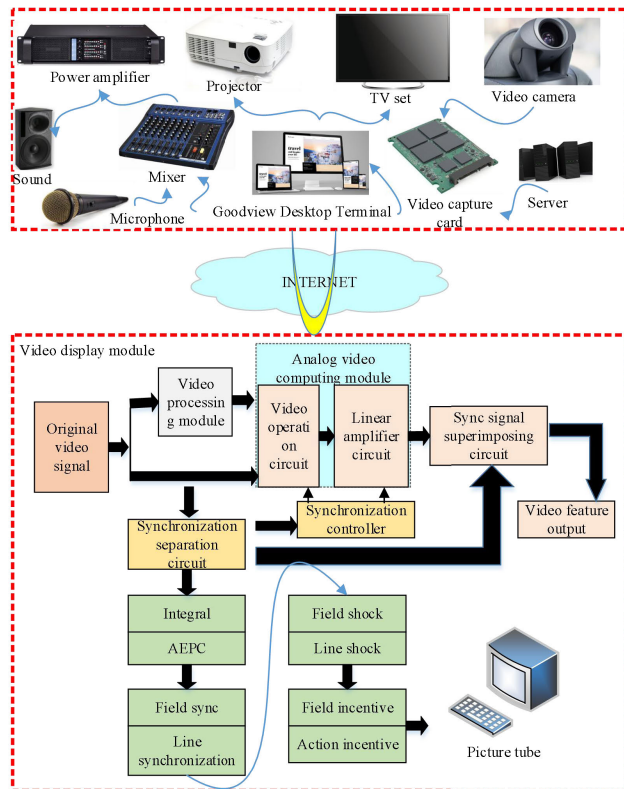

FIGURE 1. A framework for feature extraction based on mobile terminal video images.

B. SIFT DESCRIPTION

SIFT features are based on the extrinsic view of the object of interest, independent of the size and rotation of the image [23], [24]. It also has a high tolerance for light, noise, and micro-angle changes. Based on these characteristics, they stand out and are relatively easy to capture. In a large database of features, objects are easily identified and rarely misidentified. The detection rate of partially obscured targets using

SIFT feature descriptions is also high, and even more than three SIFT target features are sufficient to calculate location and orientation. Under the condition of fast hardware speed and small feature database, the recognition speed can be calculated in real time. SIFT features have a large amount of information, which is suitable for fast and accurate matching in massive databases. Figure 2 is a diagram of the SIFT feature extraction process.
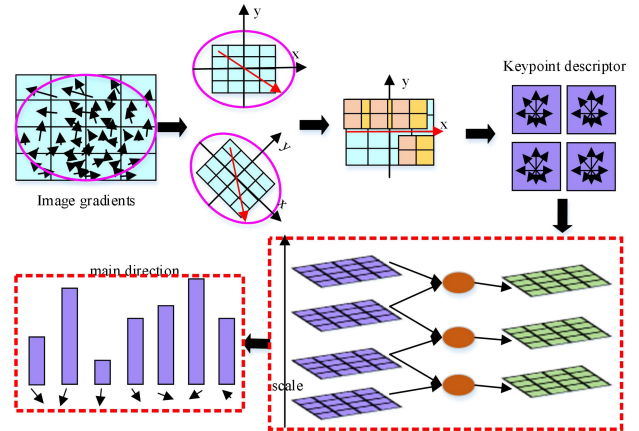

FIGURE 2. SIFT feature extraction map.

SIFT feature detection mainly includes the following four steps [25], [26]:

(1) Extremum detection in scale space
(2) Positioning of key points
(3) Direction determination
(4) Description of key points

C. SURF DESCRIPTION

The Harris feature point detection is sensitive to rotation, scale, and illumination [27]. SIFT feature extraction method has the advantages of good robustness, rotation transform, scale transform and illumination invariant. However, disadvantage of the SIFT feature extraction method is that it is slow and not suitable for real-time processing of moving target images. Compared with SIFT feature extraction, SURF feature point detection method maintains the accuracy and robustness of SIFT feature point extraction, improves the speed, and is widely used in feature extraction. Therefore, this paper adopts the SURF feature point extraction method [28].

The Hessian matrix is as follows:

$$Hessian = \begin{pmatrix} \partial_{xx}(x, \sigma) \partial_{xy}(x, \sigma) \\ \partial_{xy}(x, \sigma) \partial_{yy}(x, \sigma) \end{pmatrix} \quad (3)$$

Among them, the elements of the Hessian matrix are the partial derivatives of x and y respectively. Set the weight value of 0.9 to calculate the value of the determinant of the Hessian matrix, namely:

$$|Hessian_{approx}| = \partial_{xx}(x, \sigma) \partial_{yy}(x, \sigma) - (0.9 \partial_{xy}(x, \sigma))^2 \quad (4)$$

This is taken based on experience, and setting the weight value of 0.9 to calculate the value of the Hessian matrix

determinant can well maintain the gradient characteristics of the image.

The size of the determinant value reflects the significance of the feature in the image, and a threshold can be set to filter out points less than the threshold. Prior to this, multiple scales of Gaussian filtering should be performed on the input image, which is, corresponding to different degrees of blur, to build an image pyramid. Similar to the non-maximum suppression in the SIFT algorithm, the SURF descriptor also detects feature points through non-maximum suppression. Next, we will describe the feature points to facilitate subsequent registration and other processing. To process each detected feature point, first take the current pixel as the center, establish an N×N window in the neighborhood, and detect the main direction of each pixel. Secondly, the window is divided into multiple subwindows, and the descriptor of the SURF feature is obtained. The schematic diagram of this process is shown in Figure 3.
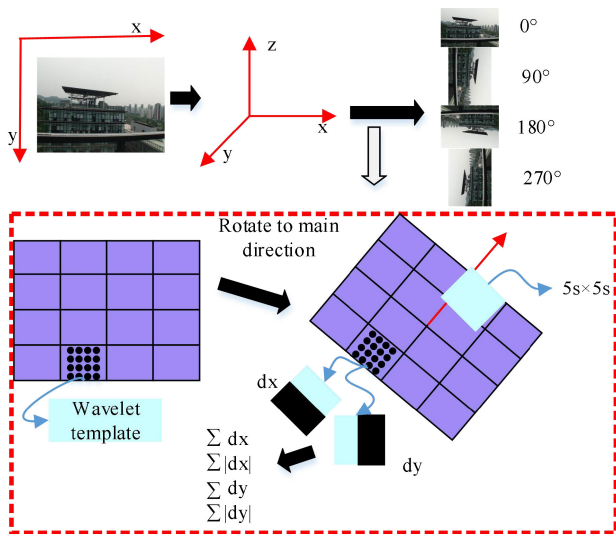

FIGURE 3. SURF feature extraction map.

III. A SURF-BASED VIDEO IMAGE FUZZY FEATURE EXTRACTION ALGORITHM

This paper proposes a method for extracting fuzzy features of mobile terminal video images based on SURF-based virtual reality technology. First, perform video image grayscale extraction on the input mobile terminal video image, and detect the closed area in the mobile terminal video image as the radiation invariant area of the terminal video image. Secondly, the Hessian matrix and the SURF description operator are used for matching to obtain the initial matching point pair. Finally, the obtained fuzzy feature one-way matching result of the video image is matched in two directions, thereby completing the mobile terminal video image fuzzy feature extraction. Figure 4 shows the algorithm flow chart of this article.

A. RADIATION-INVARIANT CLOSED AREA EXTRACTION OF MOBILE TERMINAL IMAGES

Firstly, gray histogram equalization method is used to extract the input mobile terminal video image, which emphasizes the
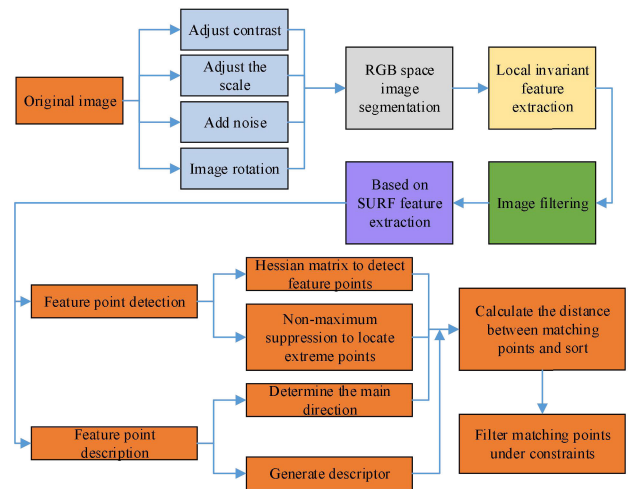

FIGURE 4. The algorithm flow chart of this article.

contrast of the mobile terminal video image and highlights the mobile terminal video image. The fuzzy feature of the edge of the target area converts the color into a grayscale video image. Secondly, after binarizing the video image of the mobile terminal, the closed area and the non-conforming area are extracted, thereby determining the closed area where the mobile terminal video image does not radiate. The specific process is as follows:

First, convert the color video image of the mobile terminal into a gray-scale video image, define the color original video image of the mobile terminal as IF, and IG as the gray-scale video image of the mobile terminal.

$$I_G = (0.299\,0.587\,0.114) \times (I_{FR}I_{FG}I_{FB})^r \qquad (5)$$

In the formula, variable I_{FR}, variable I_{FG}, variable I_{FB} represent the R, G, and B components. Probability of appearance of pixel with gray level i in the video image of the mobile terminal can be expressed as:

$$pro = \frac{Num_i}{Num}, \quad i \in [0, 1, 2, \ldots, N-1] \qquad (6)$$

In the formula, the variable N represents the number of all gray levels in the mobile terminal's gray-scale video image, and the variable Num represents the number of all pixels in the mobile terminal's video image.

Assuming that the variable A represents the cumulative probability distribution function corresponding to the variable pro, you can use the following formula to define A:

$$A(i) = \sum_{j=0}^{i} pro_i(j) = \sum_{j=0}^{i} \frac{Num_i}{Num} \qquad (7)$$

For the mobile terminal gray-scale video image IG, the gray-scale range of the mobile terminal video image after gray-scale processing is in the interval $I_{G\min} \sim I_{G\max}$, and the output video image Ig after the histogram equalization of IG according to the cumulative probability distribution

function can be expressed for:

$$I_g = I_{G\min} + (I_{G\min} - I_{G\max})A(i) \qquad (8)$$

After equalization, the gray distribution of the mobile terminal video image is relatively uniform, and the histogram corresponding to the mobile terminal video image also shows a relatively regular trend of change. After the mobile terminal video image extraction process, most of the interference area on the original terminal video image is removed, and the effect of noise is effectively reduced to a certain extent. Applying different thresholds t_i to the mobile terminal video image Ig can obtain multiple binarized mobile terminal video images Ib_i. The conversion formula from the mobile terminal gray-scale video image Ig to the binarized video image Ib_i can be defined as:

$$Ib_i = \begin{cases} 1 & \text{if g(x,y)} \geq t_i \\ 0 & \text{else} \end{cases} \qquad (9)$$

In the formula, the variable t_i determines the number of regions in Ib_i. Different gray levels can be selected to obtain different mobile terminal video image blocks Ig(x, y). The number of closed regions in the corresponding terminal video images is also different, and the radiation invariant region can be extracted from Ib_i.

Remove the range connected with the mobile terminal video image frame in Ib'_i, that is, remove the area smaller than δ, and mark the remaining mobile terminal video image as Ib'_i. Obviously the number of regions in Ib'_i is less than bi, which defines m regions in Ib'_i. By calculating the area $area_j$ of the $j - th$ area, the total area of the mobile terminal area is area, which is expressed as:

$$area = area_1 + \ldots + area_m, \quad j = 1, 2, \ldots, m \qquad (10)$$

In the formula, the variable area represents the area of the $m - th$ area. Set $Weight_i''$ to be the area weight of $i'' - th$ area of the video image of the mobile terminal, then the variable $Weight_i''$ can be defined as:

$$Weight_i'' = \frac{area_i}{area} \qquad (11)$$

The average area avg_area of m areas can be calculated according to the area weight of each area of the video image of the mobile terminal. Use the following formula to express

$$avg\_area = Weight_j area_j = \frac{area_j^2}{area} \qquad (12)$$

In the formula, the variable $Weight_j$ is the area weight of the $j - th$ area of the video image of the mobile terminal. Variable δ value is calculated for the remaining part, and the noise less than δ or the closed area with less area are eliminated.

B. SURF FEATURE POINT DETECTION

The Hessian detection operator is used to detect the fuzzy feature points in the radiative closed region of mobile terminal video images, and generate the SURF fuzzy feature description vector. Secondly, the initial matching point pair

of the video image of the mobile terminal is obtained through the fast approximate nearest neighbor search algorithm. Then, two-way matching is performed on the obtained one-way matching result of the fuzzy feature of the mobile terminal video image. Finally, the method of counting the distance mean error between fuzzy feature points is used to eliminate mismatched points.

Feature extraction and matching relationship whether feature matching can reflect the true position of a point in space. Extraction of feature points is related to the characteristics of the obtained image and also related to the matching method of feature points. Commonly used feature extraction mainly includes corner features, line features, local area features, invariance features, etc. [29]–[31]. Considering the influence of complex background environment and lighting factors, this paper adopted SURF algorithm to better extract feature points in the case of light changes and image changes. The scale invariance of this algorithm is superior to Harris algorithm, and the time complexity is relatively low. In this paper, the SURF algorithm is used, which can extract feature points better when the illumination changes and the image changes to a certain extent, and the scale invariance is better than Harris, and the time complexity is lower than SIFT.

The maximum value of matrix determinant is used to describe the location of block fuzzy feature structure of mobile terminal video image. The Hessian matrix is composed of partial derivatives of functions, which can be expressed as:

$$Hessian = \begin{pmatrix} \partial^2/\partial x^2 & \partial^2/\partial x \partial y \\ \partial^2/\partial x \partial y & \partial^2/\partial y^2 \end{pmatrix} \qquad (13)$$

Use the following formula to give the Hessian matrix discriminant:

$$Hessian' = \frac{\partial^2}{\partial x^2}\frac{\partial^2}{\partial y^2} - (\frac{\partial^2}{\partial xy}) \qquad (14)$$

All fuzzy feature points of mobile terminal video images are classified by using the symbol of judgment result. According to the discriminant, positive or negative are taken to judge whether the point is an extremum point. The second order standard Gaussian function is selected as the mobile terminal video image filter, and the second order partial derivative of the mobile terminal video image is calculated by convolution between specific cores. Thus, calculate the three matrix elements on the scale Y, and calculate the matrix with the following formula

$$G(X, Y) = \frac{\partial^2}{\partial x^2}X' \qquad (15)$$

$$Hessian(X, Y) = \begin{pmatrix} \partial_{xx}(X, Y) & \partial_{xy}(X, Y) \\ \partial_{xy}(X, Y) & \partial_{yy}(X, Y) \end{pmatrix} \qquad (16)$$

In the formula, the variable X' represents the Gaussian function, and variable $\partial_{xx}(X, Y)$ represents the convolution of second derivative of Gaussian $\frac{\partial^2}{\partial x^2}X'$ and the mobile terminal video image IF at point X. Two-dimensional movement of the

mobile terminal video image in the y, and x directions after Gaussian filtering convolution of terminal video images.

The approximate value of the matrix discriminant can calculate with the following formula:

$$Hessian_{approx} = \Psi_{xx}(X, Y)\Psi_{yy}(X, Y) - (w\Psi_{xx}(X, Y))^2 \quad (17)$$

In the calculation, the weight coefficient ω is defined as 0.9. According to the Taylor expansion, we can obtain:

$$Hessian(X) = Hessian + (\partial Hessian/\partial x) \times x$$
$$+ \frac{1}{2}(\partial^2 Hessian/\partial^2 x) \times x \quad (18)$$

When $Hessian(X) = 0$ is satisfied, we can get:

$$\tilde{X} = -(\partial^2 Hessian/\partial^2 x)(\partial Hessian/\partial x) \quad (19)$$

The derivative of the function can be approximated by the difference between neighboring pixels of the video image of the mobile terminal.
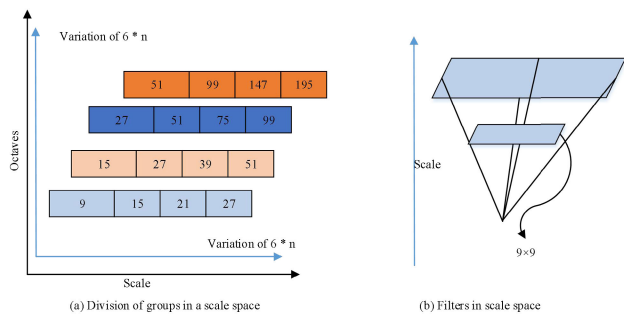

FIGURE 5. The scale space of each group.

To cover all scales, there will be overlap between groups, as shown in Figure 5(a). Figure 5(a) shows the division of each group corresponding to different scale spaces. Figure 5(b) corresponds to the 9×9 filter template in the first group of scale space, and the scale increases from the bottom to the top of the inverted pyramid.

The Hessian matrix feature point is compared with 26 other values in a $3 \times 3 \times 3$ scale space centered on the Hessian matrix. It is a local extreme point only if it is greater than or less than some other value. Then interpolate in scale space and image space to get the final feature point position and scale value. In order for the feature to have better rotation invariance, it is necessary to give each feature point a principal direction. Then, centering on the feature points, rotate the coordinate axis to the main direction, select a square area of 20s×20s (s is the proportion of the feature points), and divide it into 16 square sub-windows with each side length of 5S. Then the Gaussian weighted wavelet is obtained by using the Haar wavelet template with the scale of 2s. Finally, in each square matrix, the Haar wavelet responses in x and Y directions are summed to form a four-dimensional vector, which is normalized to form a 16×4 with a total of 64 dimensions SURF description operator.

TABLE 1. The number of SURF feature points after image rotation.

Rotating angle	0	90	180	270
Number of SURF featuring points	792	832	830	805

Table 1 compares the positions of the SURF feature points in the original image obtained after rotating the image by 0°, 90°, 180°, and 270°. Target feature points extracted after rotation and the target feature points extracted without rotation are intuitively almost overlapped, and the numbers are the same. It can be seen that the SURF description operator not only has scale and rotation invariance, but also has the same feature vector when there is no noise interference.

C. MATCHING OF FUZZY FEATURE POINTS IN VIDEO IMAGES OF MOBILE TERMINALS

The accuracy of feature point extraction directly affects the matching result. On the other hand, in image matching, due to the influence of many factors such as distortion and feature inconsistency, mismatches may occur in the matching process. In order to weaken the mismatch, the corresponding measurement method needs to be selected. This article uses similarity measurement method. Let n1 and n2 be the number of feature points of images A and B respectively, and variable A_j and variable B_j are any feature points of images A and B respectively, then the distance similarity between variable A_j and variable B_j is equation (20):

$$Similar(A_j, B_j) = \sum_{j=0}^{m} (A_j, B_j)^2 \quad (20)$$

In formula, the variable m is the dimension. The smaller the variable $Similar(A_j, B_j)$, the higher the similarity of the feature points, but this is likely to cause two mismatches.

For this reason, this paper uses nearest distance to next nearest distance to find the most obvious feature points within the set threshold. The method is defined as: Let $Similar1$ be shortest distance (dimensionless), $Similar2$ be the sub-close distance, η is the set threshold ($0 < \eta \leq 1$), and must satisfy condition shown in equation (21).

In this paper, local epiploic constraints, left and right coordinate constraints, unique constraints, are integrated to improve the correct matching rate.

$$ratio = Similar1/Similar2 \leq \eta else$$
$$ratio = Similar1/Similar2 > \eta \quad (21)$$

In practical applications, the x-axis of parallel binocular camera is not collinear, and there is a slight deviation. Therefore, the epiploic constraint cannot be directly used when matching the left and right target points. Therefore, local limit constraints are proposed. That is, there is a left and right target point matching search area on longitudinal y axis, as shown

in equation (22).

$$|y_{left} - y_{right}| < threshold \qquad (22)$$

Among them, variable y_{left} and variable y_{right} are the ordinates of the left and right matching points respectively. If the value of threshold is too small, it will degenerate into a polar constraint. This causes a large number of correct matching point pairs to be filtered out; if the threshold value is too large, incorrect matching point pairs cannot be filtered out, and it will increase time consumption.

The left and right coordinate constraint is that the x coordinate of the target point in the left image is greater than the corresponding x coordinate in the right image [32], and the difference can be expressed as equation (23).

$$y_{left} - y_{right} = \frac{ZV}{Ls} \qquad (23)$$

Among them, the variable L is the z coordinate of the target point in the camera coordinate system, the variable Z is the length of the baseline, the variable V is the focal length, the variable s is the physical size of the pixel, and the variable y_{left} is the target point The x coordinate in the left image, and the variable y_{right} is the corresponding x coordinate of the target point in the right image. The physical quantities on the right side of equation (23) are all positive. Therefore, the formula (24) is established, which is the coordinate constraint of the left and right target points.

$$y_{left} > y_{right} \qquad (24)$$

The principle of uniqueness constraint: a feature point in the left image, if a matching point exists in the right image, it is unique. Combining formulas (22), (24) and uniqueness constraints, the matching area will be further reduced. Obviously, this greatly reduces the area of the target point matching search.

D. IMAGE FEATURE RECOGNITION BASED ON HMM MODEL

The hidden Markov model is selected for feature classification, and the more obvious boundary features in the image are selected as image features. Aiming at the image boundary, a method based on the Euclidean distance from the image centroid to the boundary is proposed, which divides the boundary from 0 to 180 degrees into 90-dimensional features. In this process, the centroid coordinates need to be calculated, as shown in the following formula:

$$P_c^* = \frac{\sum_{ij} P_i F_{ij} W_{ij}}{\sum_{ij} F_{ij} W_{ij}} \qquad (25)$$

$$Q_c^* = \frac{\sum_{ij} Q_i F_{ij} W_{ij}}{\sum_{ij} F_{ij} W_{ij}} \qquad (26)$$

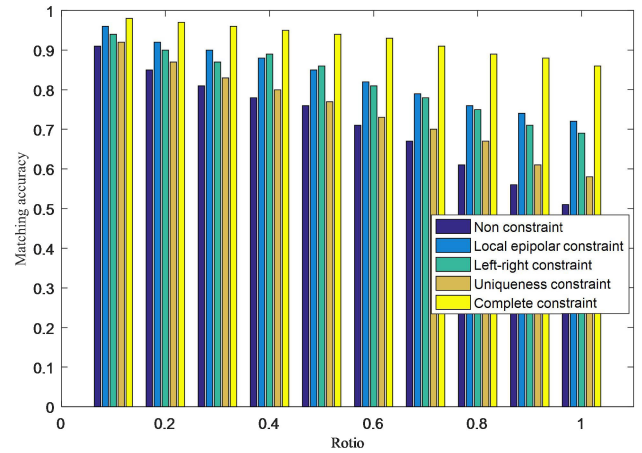$$W_{ij} = \frac{1}{\sqrt{(P_i - P_c)^2 - (Q_i - Q_c)^2}} \qquad (27)$$

FIGURE 6. Comparison of correct matching rate with different ratios under different constraints.

The HMM model is selected to classify the image features. The HMM model is a collection of transferred states. Each step of the transition corresponds to the transition probability and the output probability. The output probability is the probability of forming an output value from a large number of values in a specific state. The transition probability refers to the probability of each state transitioning to each other.

Set the HMM model as:

The number of N state sets is $\{R_1, R_2, \ldots, R_N\}$, and the state at time u is expressed by a random variable H_u. The set of M intuitive values is $\{r_1, r_2, \ldots, r_N\}$, the intuitive sequence at time u is expressed by the random variable J_u, and the intuitive sequence corresponds to the output value of the HMM obvious state variables.

An N×N state transition movement probability distribution matrix is $P = \{O_{ij}\}$, and the probability expression is:

$$O_{ij} = P(H_{u+1} = R_j | H_{u+1} = R_i) * (P_c^*, Q_c^*) \qquad (28)$$

An intuitive probability distribution matrix is $S = \{O_j(k)\}$, and the variable $O_j(k)$ represents the probability of outputting $O_j(k)$ in the R_j state at time u:

$$O_j(k) = P(H_{u+1} = O_{ij} | H_{u+1} = R_i) * (P_c^*, Q_c^*) \qquad (29)$$

An original state distribution matrix $M = \{M_i\}$, the variable M_i is the probability that the original state is R_i:

$$M_i = P(H_{u+1} = R_i) \qquad (30)$$

In summary, the hidden Markov model can be expressed as:

$$\zeta = NM_i/AB \qquad (31)$$

The purpose of image feature recognition through the HMM model is to make the image feature recognition more accurate, while also reducing the time consumed by redundant steps and improving the computational efficiency.
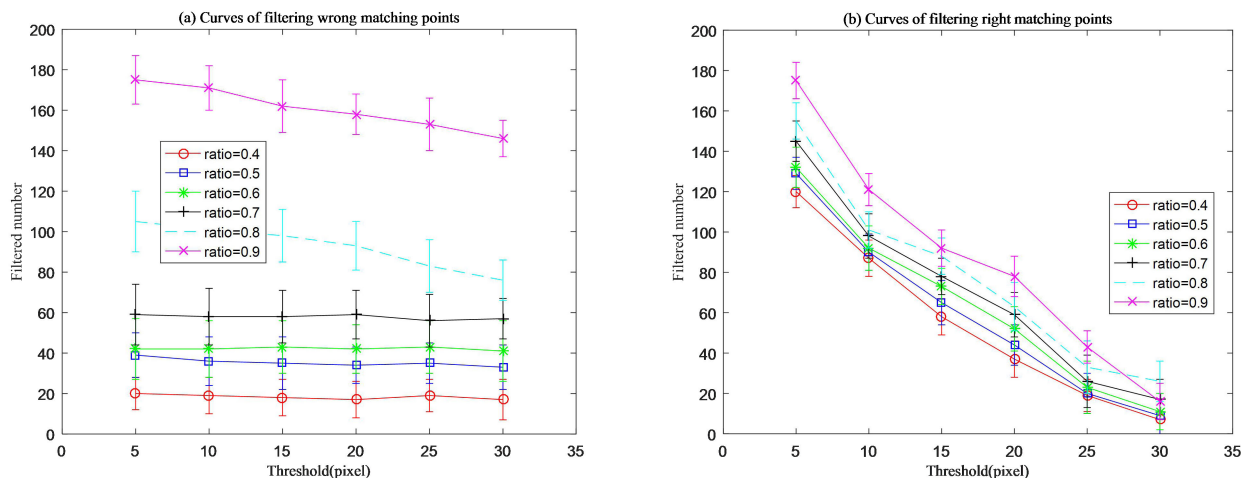
FIGURE 7. Use different ratio matching points to filter out statistical comparison.

IV. EXPERIMENTAL VERIFICATION

A. EXPERIMENTAL DATA DESCRIPTION

The experiment was carried out on the virtual machine of ARM CPU, 343M, and Android 4.4 system in Eclipse software.

The experimental data selects a total of 500 pictures collected on the Internet and taken by actual mobile terminals, which are used as experimental reference sample data, covering the multi-view, and multi-color changes of the mobile device terminal. 300 pictures formed by transformations such as rotation, brightness, adding Gaussian white noise and 50 pictures taken by the actual mobile terminal are used as the test set, which are divided into two video image set types A and B according to the video image size type. As shown in Table 2, simulation parameters are shown.

TABLE 2. Simulation parameters.

Image group	Pixel values	Size	Ratio
A	1.6 million	1200*800	16：9
B	1.2 million	400* 300	4：3

B. THRESHOLD SELECTION FOR FEATURE POINT MATCHING

When the ratio increases, the correct matching rate decreases faster, and the result is shown by the dotted line in Figure 6. Therefore, it is necessary to further filter out incorrect matching points. Figure 6 shows the statistical results of comparing the correct matching rate with the above three constraints and full constraints under different ratio conditions. It can be seen that in this experiment, the correct matching rate using local constraints, left and right coordinate constraints, and uniqueness constraints is sequentially reduced. At the same time, the correct matching rate using full constraints does not change much within the range of ratio, and the correct rate can still reach 90% under full constraints. On the other hand, considering that the ratio is smaller, the number of detected feature points will be reduced, so in order to ensure

TABLE 3. Comparison of running time of different methods for terminal video image.

Terminal video image data set	Method of this article	Literature [19]	Literature [20]
A	0.551	2.671	3.213
B	0.667	3.761	3.911

a sufficient number of feature points, the ratio needs to be as large as possible. Therefore, ratio = 0.9 is more appropriate in this experiment.

As shown in Figure 7, it is a comparative analysis of filtering out wrong matching points and filtering out correct matching points under different ratio conditions for images. Obviously, the filtering curve represented by different ratios in Figure 7(a) is close to the level when the local constraint range is 5-30, indicating that the mismatching has no effect on filtering in this local range. Figure 7(b) shows the change curve that uses different ratios to filter out the correct matching points within the local constraint range. Obviously, the curve shows a downward trend, indicating that the correct matching points filtered out within the local constraint range gradually decrease, and it is close to 0 at 20. In summary, it is more appropriate to take the threshold of the local constraint range threshold = 20 in this experiment.

C. ANALYSIS OF THE RESULTS OF DIFFERENT ALGORITHMS FOR IMAGE FUZZY FEATURE EXTRACTION

In this paper, fuzzy feature extraction experiments are carried out on mobile terminal video images using the method presented in this paper, the method in literature [19] and the method in literature [20]. The mobile terminal video image A and B are used to extract the fuzzy feature. Specific results are shown in Table 3.

Table 3 shows that with the gradual increase in the size of mobile terminal video images, the time required to extract fuzzy features of mobile terminal video images using the

method in this paper is much shorter than that in the literature [19] and the literature [20] method, which can fully satisfy the virtual reality technology under the real-time requirements of mobile terminals.

The paper, literature [19], and literature [20] were used to carry out fuzzy feature extraction experiments of mobile terminal video images. In order to verify the extraction effects of different methods on fuzzy features of mobile terminal video images, Figure 8 shows the extraction effects of different methods. As can be seen from Figure 8, the fuzzy features of the mobile video image are shown in Figure 8(a), and there are many fuzzy features. As can be seen from the experimental results, Figure 8(b) shows the effect of the fuzzy feature extraction method in reference [19]. Figure 8(c) is the result of the extraction method in this paper. The contrast of local texture details of mobile terminal video image increases, which is obviously better than the method mentioned in the literature.

D. ANALYSIS OF THE RESULTS OF DIFFERENT ALGORITHMS FOR FEATURE MATCHING POINTS

Figure 9 shows two test images using different methods for feature extraction and matching experiments. Figure 9(a) and Figure 9(b) are the test images using SIFT algorithm to extract image features. Figure9(c) is the result of matching the image features extracted by SIFT. Figures 9(d) and 9(e) are the experimental results obtained by the SURF feature extraction algorithm used in this paper, and Figure 9(f) is the result of matching the features extracted by the SURF algorithm. From the perspective of visual effects, these figures all extract the feature information of the two images to varying degrees, but the feature information extracted by SURF in this article with more points, the amount of information expressed is richer. Figure 9(c) and Figure 9 (f) after the feature information is quickly matched, the matching speed and success rate are also ideal.

In this paper, four indicators of SIFT algorithm and SURF algorithm were counted. These indexes are the number of feature points, the number of matching points, the matching success rate, and the running time respectively. In this paper, these indexes are counted and the quality of two different algorithms is evaluated. The number of feature points reflects the feature extraction capability of the algorithm. The more feature points, the more detailed image information. The number of matching points and the power of matching results reflect the quality of image matching. The running time reflects the efficiency of the algorithm. Table 4 shows that method proposed by paper has more feature points and matching points, higher matching success rate, and shorter running time.

Using the histogram affine method, Harris method, SIFT method, literature [19], literature [20] and the proposed method to carry out two sets of comparison experiments for the repeated feature rate, and the comparison result is shown in Figure 10.

(a)Original image

(b)Literature[19] extract features

(c)This paper method extract features
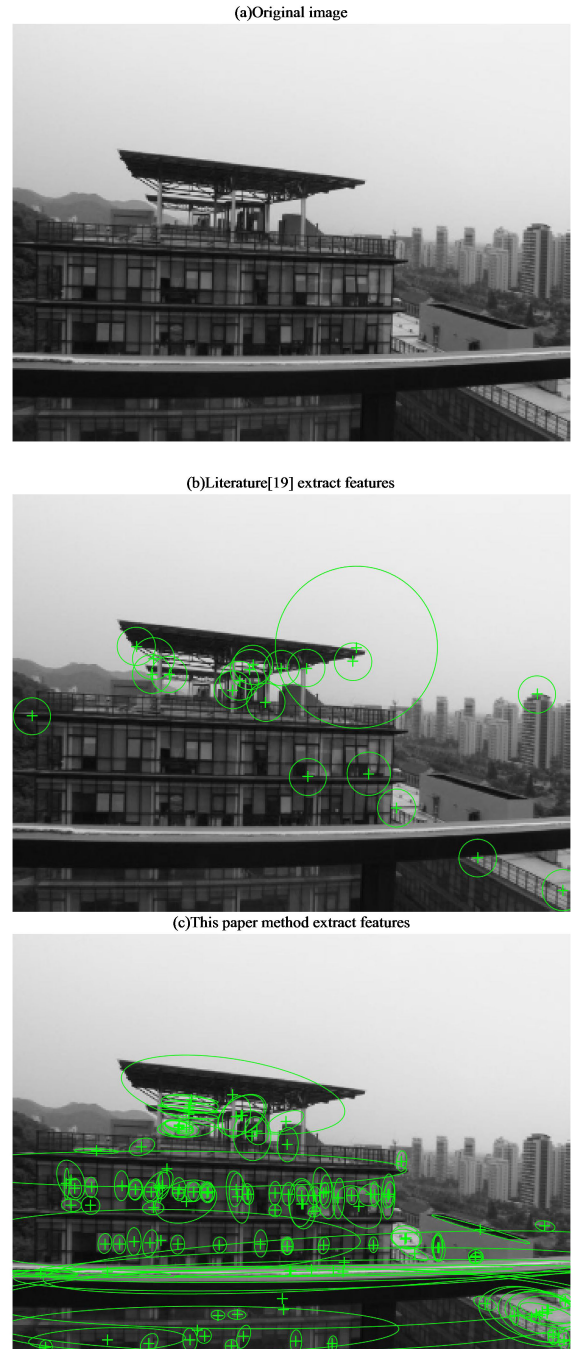
FIGURE 8. Different methods of video image fuzzy feature extraction effect.

TABLE 4. Comparison of SIFT and SURF feature matching data.

Feature extraction method	Feature points	Number of matching points	Match success rate	Operation hours
SIFT	107	78	0.728	1.012
SURF	124	97	0.782	0.881

Analyzing the experimental results of Figure 10, it can be seen that Figure 10(a) is the result of calculating the feature repetition rate of group A, and Figure 10(b) is the result of
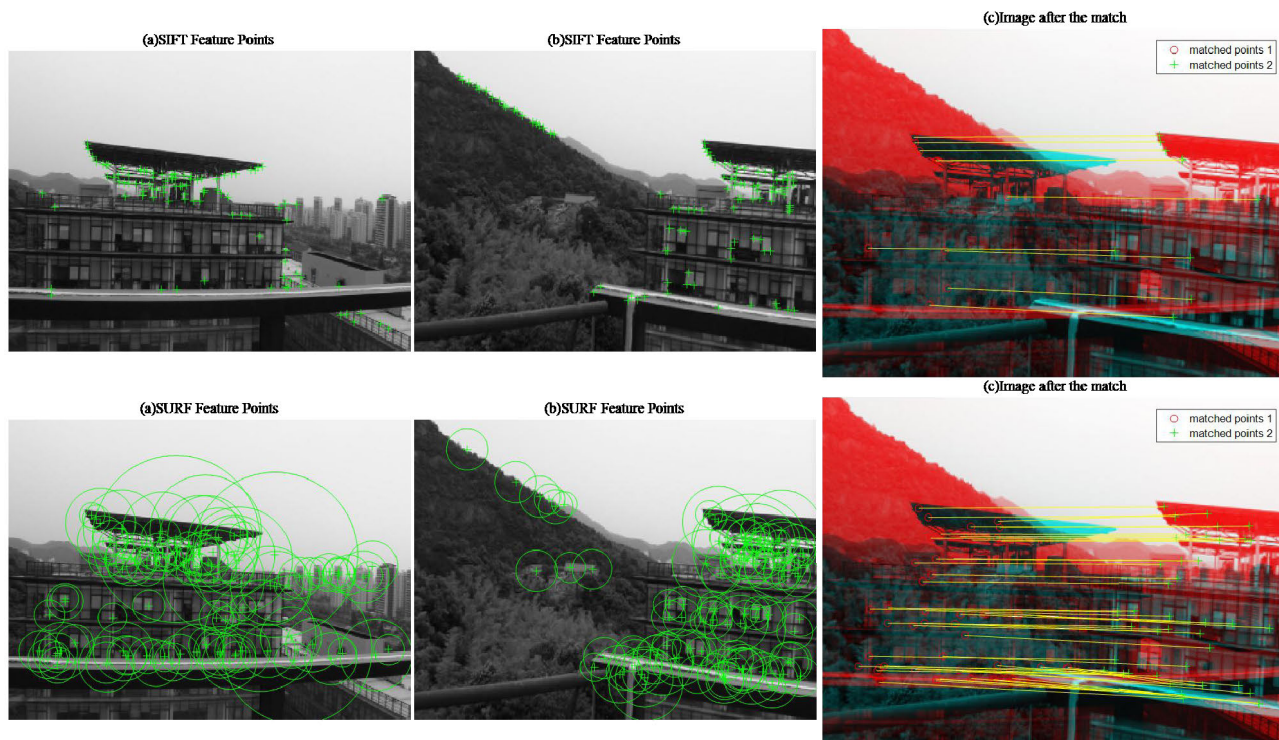
FIGURE 9. SIFT and SURF features are matched and compared.

calculating feature repetition rate of group B. Combining the two sets of image feature repetition rates, for images with different pixels, the larger the pixel, the greater the number of features, and the higher the repetitive feature rate. It can be seen that features with low pixels are less robust in recognition. The histogram affine method used in the A group of images to identify the repetitive feature rate is higher, but in the B group of images, the feature repetition rate recognized by this method is lower, indicating that this method is not suitable for images with higher pixels. For group A, the Harris method has a poor recognition effect. In group B, although the highest value of the recognition feature repetition rate is between 0.75% and 0.80%, the recognition effect of this method fluctuates. From an intuitive point of view, the trend of broken line change is too frequent, indicating that the recognition effect of this method is unstable. However, the proposed method performs well in both sets of experiments. It can be seen from the figure that the variation of the feature repetition rate recognized by the proposed method is small and the change is relatively stable, indicating that the proposed method is more robust and the recognition effect is excellent and feasible.

E. ANALYSIS OF RECOGNITION RESULTS AFTER DIFFERENT ALGORITHM FEATURE EXTRACTION

Perform the next experiment on the taken experimental images, using histogram affine method, Harris method, SIFT method, literature [19], literature [20] to compare with the proposed method to verify the credibility of the proposed

TABLE 5. Image feature recognition results.

	Identification number	Misunderstanding rate	Time
Histogram affine method	7	0.307	1.612
Harris method	5	0.529	1.241
SIFT method	4	0.412	1.452
Literature [19]	4	0.312	1.987
Literature [20]	5	0.199	2.165
This paper method	9	0.101	0.912

method. The purpose of preprocessing the collected experimental objects is to improve the identification efficiency. First select an image, convert the image to a grayscale image, and perform binarization processing on the grayscale image to obtain an image to be recognized with a size of 32*32. The histogram affine method, Harris method and the proposed method are used to perform image feature recognition experiments. The experimental comparison results are shown in Table 5 and Figure 11.

Experimental results show that number of recognition of the histogram affine method is 7, which is higher among the three methods, and the misrecognition rate is relatively low, which has a better recognition effect, but the time used
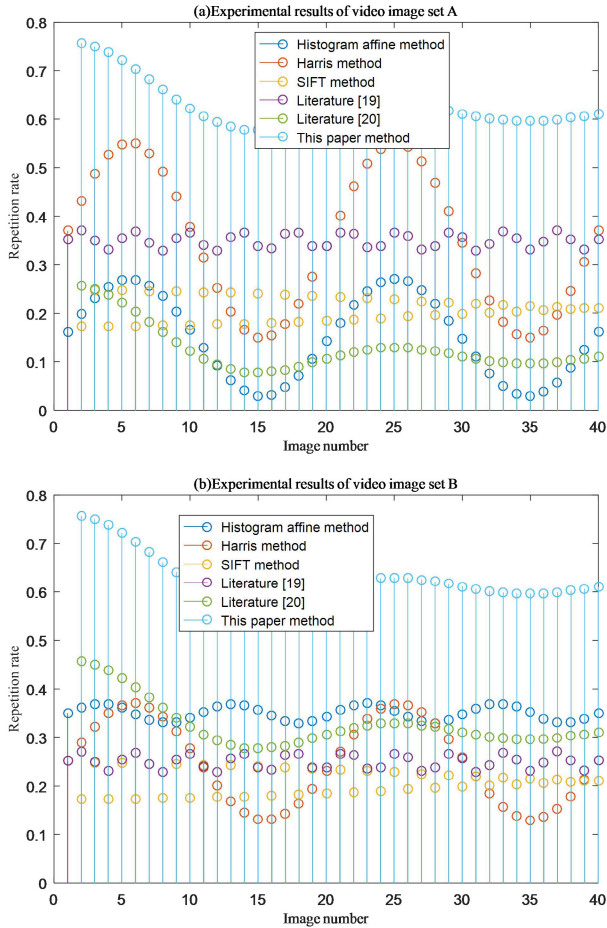
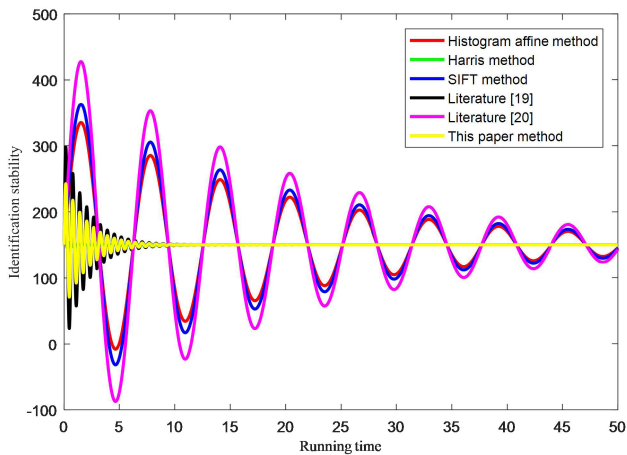**FIGURE 10. Experimental results of two sets of images.**

FIGURE 11. Comparison of recognition stability of different methods.

by this method is three The most of the methods indicates that the operating speed of the identification method is low. Although the Harris method has shorter operation time than the histogram affine method, the number of recognition and the misrecognition rate are the lowest, and there is a problem of lower recognition accuracy. The method proposed by paper has the highest recognition times and the lowest

error recognition rate, indicating that this method has a high recognition accuracy rate of 0.41s, which meets the requirements for computing speed. According to the analysis of the experimental results, it can be proved that the proposed method has a better recognition effect and is short in time.

V. CONCLUSION

Mobile terminals can be used to realize voice feature recognition, image feature recognition and other operations. Most mobile devices operate directly through gesture commands. As the most mainstream technology of mobile terminals, image feature recognition technology mainly uses the high-speed processing capabilities of computers to help people process large amounts of information and recognize various targets.

According to the actual requirements for image feature recognition of mobile devices and the existing problems in existing image feature recognition methods, a method for extracting video image features of mobile terminals based on SURF virtual reality technology is proposed. Through in-depth research on the reality technology of mobile terminals, we selected sample video images that have obvious differences in color, light, angle, etc., and included a variety of mobile devices, and performed video image blur feature extraction tests on two smart phones. First, perform video image grayscale extraction on the input mobile terminal video image, and detect the closed area in the mobile terminal video image as the radiation invariant area of the terminal video image. Secondly, Hessian matrix and SURF description operator are used for matching to obtain the initial matching point pair. Finally, the obtained fuzzy feature one-way matching result of the video image is matched in two directions. Results show that the proposed method has higher computational efficiency and better recognition effect, can meet the actual requirements of image feature recognition, and has practical application advantages.

REFERENCES

[1] C. Donghui, L. Guanfa, Z. Wensheng, L. Qiyuan, B. Shuping, and L. Xiaokang, "Virtual reality technology applied in digitalization of cultural heritage," *Cluster Comput.*, vol. 22, no. 4, pp. 10063–10074, Aug. 2019.

[2] A. Schnack, M. J. Wright, and J. L. Holdershaw, "Immersive virtual reality technology in a three-dimensional virtual simulated store: Investigating telepresence and usability," *Food Res. Int.*, vol. 117, pp. 40–49, Mar. 2019.

[3] D. M. Shafer, C. P. Carbonara, and M. F. Korpi, "Factors affecting enjoyment of virtual reality games: A comparison involving consumer-grade virtual reality technology," *Games Health J.*, vol. 8, no. 1, pp. 15–23, Feb. 2019.

[4] J. Lin, L. Cao, and N. Li, "Assessing the influence of repeated exposures and mental stress on human wayfinding performance in indoor environments using virtual reality technology," *Adv. Eng. Informat.*, vol. 39, pp. 53–61, Jan. 2019.

[5] K. Sui and H.-G. Kim, "Research on application of multimedia image processing technology based on wavelet transform," *EURASIP J. Image Video Process.*, vol. 2019, no. 1, pp. 1–9, Jan. 2019.

[6] L. Liu, J. Xu, Y. Huan, Z. Zou, S.-C. Yeh, and L.-R. Zheng, "A smart dental health-IoT platform based on intelligent hardware, deep learning, and mobile terminal," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 3, pp. 898–906, Mar. 2020.

[7] H. Jiang, "Mobile fire evacuation system for large public buildings based on artificial intelligence and IoT," *IEEE Access*, vol. 7, pp. 64101–64109, May 2019.

[8] P. Wu, W. Li, and W. Song, "Fast, accurate normalized cross-correlation image matching," *J. Intell. Fuzzy Syst.*, vol. 37, no. 4, pp. 4431–4436, Oct. 2019.

[9] X. Wei, Z. Yang, Y. Liu, D. Wei, L. Jia, and Y. Li, "Railway track fastener defect detection based on image processing and deep learning techniques: A comparative study," *Eng. Appl. Artif. Intell.*, vol. 80, pp. 66–81, Apr. 2019.

[10] M. Y. Valandar, M. J. Barani, P. Ayubi, and M. Aghazadeh, "An integer wavelet transform image steganography method based on 3D sine chaotic map," *Multimedia Tools Appl.*, vol. 78, no. 8, pp. 9971–9989, Apr. 2019.

[11] Z. Tang, Z. Ding, R. Zeng, Y. Wang, J. Wen, L. Bian, and C. Yang, "Multi-threshold corner detection and region matching algorithm based on texture classification," *IEEE Access*, vol. 7, pp. 128372–128383, Sep. 2019.

[12] N. A. Seada and M. G. M. Mostafa, "Automatic ostia detection in CTA volume data: A comparative study," *Int. J. Med. Eng. Inform.*, vol. 11, no. 1, pp. 86–101, Dec. 2019.

[13] E. Cuevas, M.-A. Díaz-Cortes, and E. Mezura-Montes, "Corner detection of intensity images with cellular neural networks (CNN) and evolutionary techniques," *Neurocomputing*, vol. 347, pp. 82–93, Jun. 2019.

[14] B. Wang and R. Tian, "Judgement of critical state of water film rupture on corrugated plate wall based on SIFT feature selection algorithm and SVM classification method," *Nucl. Eng. Des.*, vol. 347, pp. 132–139, Jun. 2019.

[15] R. Wang, Y. Shi, and W. Cao, "GA-SURF: A new speeded-up robust feature extraction algorithm for multispectral images based on geometric algebra," *Pattern Recognit. Lett.*, vol. 127, pp. 11–17, Nov. 2019.

[16] Q. He, B. He, Y. Zhang, and H. Fang, "Multimedia based fast face recognition algorithm of speed up robust features," *Multimedia Tools Appl.*, vol. 78, no. 17, pp. 24035–24045, Jan. 2019.

[17] M. Sajjad, M. Nasir, K. Muhammad, S. Khan, Z. Jan, A. K. Sangaiah, M. Elhoseny, and S. W. Baik, "Raspberry pi assisted face recognition framework for enhanced law-enforcement services in smart cities," *Future Gener. Comput. Syst.*, vol. 108, pp. 995–1007, Jul. 2020.

[18] X. Li, H. Jiao, and Y. Wang, "Edge detection algorithm of cancer image based on deep learning," *Bioengineered*, vol. 11, no. 1, pp. 693–707, Jan. 2020.

[19] M. Hashemzadeh and A. Zademehdi, "Fire detection for video surveillance applications using ICA K-medoids-based color model and efficient spatio-temporal visual features," *Expert Syst. Appl.*, vol. 130, pp. 60–78, Sep. 2019.

[20] M. Al-wswasi and A. Ivanov, "A novel and smart interactive feature recognition system for rotational parts using a STEP file," *Int. J. Adv. Manuf. Technol.*, vol. 104, nos. 1–4, pp. 261–284, May 2019.

[21] H. Du, G. Li, S. Wang, and F. Zhang, "Discriminant locality preserving projections based on l2,-norm for image feature extraction and recognition," *J. Vis. Commun. Image Represent.*, vol. 58, pp. 166–177, Jan. 2019.

[22] A. Manickam, E. Devarasan, G. Manogaran, M. K. Priyan, R. Varatharajan, C.-H. Hsu, and R. Krishnamoorthi, "Score level based latent fingerprint enhancement and matching using SIFT feature," *Multimedia Tools Appl.*, vol. 78, no. 3, pp. 3065–3085, Feb. 2019.

[23] H.-H. Chang, G.-L. Wu, and M.-H. Chiang, "Remote sensing image registration based on modified SIFT and feature slope grouping," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 9, pp. 1363–1367, Sep. 2019.

[24] Q. Yu, S. Zhou, Y. Jiang, P. Wu, and Y. Xu, "High-performance SAR image matching using improved SIFT framework based on rolling guidance filter and ROEWA-powered feature," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 3, pp. 920–933, Mar. 2019.

[25] P. Chhabra, N. K. Garg, and M. Kumar, "Content-based image retrieval system using ORB and SIFT features," *Neural Comput. Appl.*, vol. 32, no. 7, pp. 2725–2733, Apr. 2020.

[26] G. Doménech-Asensi, J. Zapata-Pérez, R. Ruiz-Merino, J. A. López-Alcantud, J. Á. Díaz-Madrid, V. M. Brea, and P. López, "All-hardware SIFT implementation for real-time VGA images feature extraction," *J. Real-Time Image Process.*, vol. 17, no. 2, pp. 371–382, Apr. 2020.

[27] P. Sathiya and P. Anandhakumar, "Probabilistic collision estimation for tracked vehicles based on corner point self-activation approach," *Comput. Electr. Eng.*, vol. 74, pp. 557–568, Mar. 2019.

[28] F. Baig, Z. Mehmood, M. Rashid, M. A. Javid, A. Rehman, T. Saba, and A. Adnan, "Boosting the performance of the BoVW model using SURF–CoHOG-Based sparse features with relevance feedback for CBIR," *Iranian J. Sci. Technol., Trans. Electr. Eng.*, vol. 44, no. 1, pp. 99–118, Mar. 2020.

[29] J. Li, Q. Hu, and M. Ai, "RIFT: Multi-modal image matching based on radiation-variation insensitive feature transform," *IEEE Trans. Image Process.*, vol. 29, pp. 3296–3310, Dec. 2019, doi: 10.1109/TIP.2019.2959244.

[30] H. Anzid, G. L. Goic, A. Bekkari, A. Mansouri, and D. Mammass, "Multimodal images classification using dense SURF, spectral information and support vector machine," *Procedia Comput. Sci.*, vol. 148, pp. 107–115, Feb. 2019.

[31] A. V. Yoganand, A. C. Kavida, and D. R. Devi, "Pose and occlusion invariant face recognition system for video surveillance using extensive feature set," *Int. J. Biomed. Eng. Technol.*, vol. 33, no. 3, pp. 222–239, Jun. 2020.

[32] K. Zhang, G. Cheng, and L. Xu, "Topology optimization considering overhang constraint in additive manufacturing," *Comput. Struct.*, vol. 212, pp. 86–100, Feb. 2019.

YINGYING FENG was born in Shandong, China, in 1982. She received the bachelor's degree from Changjiang University, in 2004, and the master's degree from the China University of Geosciences, in 2007. She currently works with Fuyang Normal University. She has published six articles. Her research interests are included computer graphics and image processing and video analysis.

• • •