

RESEARCH

Open Access



Registration of infrared and visible light image based on visual saliency and scale invariant feature transform

Gang Liu*, Zhonghua Liu, Sen Liu, Jianwei Ma and Fei Wang

Abstract

Visual saliency is a type of visual feature which simulates visual attention selection mechanism in biological system and has better robustness and invariance. A rapid infrared image and visible light image registration method based on visual saliency and SIFT (scale invariant feature transform) is proposed in this paper. The method adopts amplitude modulation Fourier transform to construct saliency map, and the image salient points are achieved preliminarily by salient threshold. Subsequently, this method calculates the local entropy for these salient points because of entropy's character for information measurement, then these points are reordered and screened based on the strategy of entropy priority. The screened results are thought as centers for salient regions. Morphological operation is used for growing and merging for neighbor salient scene region in image. Aiming at the abstracted salient region for image, PCA (principal component analysis)-SIFT algorithm is proposed, which can produce the compressed SIFT registration features and largely reduce the computational cost of image registration. The proposed algorithm adopts random sampling conformance method to remove the mistaken point pairs before calculating the model parameter of affine transformation for registration between infrared image and visible lights. The experimental results indicate that the method has good invariance under image scale, rotation, translation, and illumination variation and can realize effective registration between infrared image and visible lights. Compared with some classical algorithms, the proposed method has advantage in registration accuracy and registration speed obviously.

Keywords: Image registration, Infrared and visible light, Visual saliency, Scale invariant feature transform, Principal component analysis

1 Introduction

Image registration technique refers to the process of transforming two or more images from different sensor, viewpoint, and time in the same scene into the same coordinate system. The transforming parameters of images are determined by some similar measure. It has been widely used in military, remote sensing, medical, computer vision, and other fields. Image registration can be divided into single-mode registration and multi-modals.

The problem of single-mode image registration has been solved in the end of the twentieth century; however, multi-modal image registration is still not well

settled. Multi-modal image which is derived from different imaging equipment can provide more abundant and comprehensive information than single-mode image. As infrared imaging reflects the radiation information of scene and visible lights reflect the reflection information, so the output images have different gray level features, and these information are complementary to each other and suited to be fused together for target positioning and identification. The premise of fusion is the registration. Therefore, the registration between infrared image and visible lights is one of the most typical multi-modal image registration.

Existing registration methods for infrared image and visible lights are mainly divided into two categories: region-based approach [1–5] and feature based [6–11]. The region-based registration method mainly uses the

* Correspondence: 13849946320@139.com

Information Engineering College, Henan University of Science and Technology, No.263 Kaiyuan Avenue, Luolong District, Luoyang 471023, China

gray information in the template or some kind of transformation of the gray information. It takes the template as a unit and calculates the degree of similarity between the current window and the template at each position of the image according to some similarity measure. Due to the small correlation of gray-scale attributes of heterogeneous images, the region-based method widely used in the field of homogeneous image registration is limited in the field of heterogeneous image registration. The feature-based registration method does not directly use the gray information, which makes the method become a very popular research direction in the field of heterogeneous image matching. At present, the characteristics of the registration process are mainly focused on point features [7–13], line features [14–16], and invariant moment features [17–19]. The feature-based matching algorithm is more suitable for the case where the information of the image is rich, and the objects in the image are easy to distinguish and detect. The disadvantage is that the result of feature extraction is closely related to the content and quality of the image, and its robustness is poor. It is the most critical problem to find the characteristic description operator with good stability and robustness between multi-source sensors. At present, most algorithms can only get better results on the images chosen by the researchers.

In general, the process of image registration consists of feature space, search space, similarity measure, and search strategy. The registration efficiency can be improved from the above four factors. At present, multi-modal registration based on local invariant feature is a hotspot in the field. Local invariant feature refers to the detection of local features or description which remains unchanged to the various changes for image, such as geometric transformation, illumination transformation, convolution transformation, and viewing angle changes. The basic idea of the local invariant feature is to extract the essential attribute characteristics of the image content, which are independent of the specific manifestations of the image content or they have characteristic of adaptivity.

The local invariant feature can not only obtain reliable matching under the condition that the observation condition changes greatly, shaded and cluttered interference, but also effectively describe the image content for image retrieval or scene, target recognition, and so on. It is very difficult to segment the foreground goal from the complex background, and it is hard to realize the meaningful segmentation based on the low-level feature. The image content is expressed as a collection of local invariant regions, which can avoid the segmentation problem. The method of local invariant feature is essentially an implicit segmentation of the image content, and the local invariant feature may be in prospect target, background, or the boundary of the target. The subsequent high-level

processing need to extract the information of interest based on the local invariant feature. In the existing local feature description algorithm, the scale invariant feature transform (SIFT) feature has high uniqueness and can achieve better recognition effect [20]. It is a promising direction to solve the problem of registration between infrared and visible light image under the premise of improving the efficiency of SIFT feature extraction.

The system of human vision has the ability of filtering and screening of information. In daily life, the surrounding environment constantly transmits all sorts of change information to the human eye which far beyond the human eye processing limit. The human eye will actively select to deal with the most critical part of the visual information and the academic community names it as the visual attention mechanism. In the field of computer vision, the introduction of visual attention mechanism enable machine vision to process and distribute information resources. The visual attention mechanism has always been an interdisciplinary subject involving disciplines such as neurology, anatomy, psychology, machine vision, artificial intelligence, and information processing.

At present, visual attention mechanism has two main types of modeling methods: bottom-up and top-down respectively. Top-down attention is affected by the specific tasks and people's subjective consciousness, and the results often show great differences. In comparison to top-down, bottom-up on the exploration of the common mechanism of specific tasks and specific characters are not bounded to the human visual system and has more controlled research environment and more extensive application scenarios. A typical representative of the bottom-up approach is the visual saliency map of multi-feature and multi-scale based on the neurobiological framework which is put forward by Siagian in 2007 [21]. This approach can quantify the degree of stimulation of the human visual system in the image itself and has a scale invariance. However, the effectiveness of the proposed method is not good because the center-surround operator is adopted, and the choice of parameters directly affects the saliency map. In 2007, Hou proposes a spectral residual algorithm which has the advantages of simple structure, fast computation speed [22]. However, this method has some shortcomings in particle noise sensitivity and color image processing. In 2010, Chen improves the spectral residual method and proposes the algorithm framework of amplitude modulation Fourier transform which solves the problem of noise sensitivity of spectral residual algorithms [23].

Visual saliency is a kind of visual feature with better robustness and invariance which simulates biological vision attention selection mechanism. In the SIFT algorithm, the equivalent recognition result can be realized by using some significant feature vectors. Therefore, if these salient feature points can be found accurately, the

computation time can be effectively reduced while decreasing the mismatching rate of SIFT characteristics.

Using SIFT to describe the target characteristics, it has the invariance of translation, rotation, and contraction transformation. However, its characteristic component information has the redundancy, if it is not eliminated, the registration accuracy and real time of the registration process will be reduced. PCA (principal component analysis) is a statistical method that constructs a few main indicators from multiple indicators. The principal component analysis is applied to feature selection, which can eliminate the influence of multiple features that have little contribution to the target registration, and can construct the feature that can reflect the target effectively.

In view of this, this paper proposes an image registration algorithm which combines visual saliency and SIFT based on the optimization of search space and feature space. After extracting the salient region of image, PCA method is used to establish the dimensionality reduction of SIFT feature, and finally, the robust registration of infrared image and visible lights is realized. The contents of this paper are as follows. Section 1 gives some registration methods between infrared image and visible lights and existing problems. Section 2 utilizes the theory of salient region detection and puts forward the detection method of salient region center of image based on entropy priority strategy. On this basis, Section 3 gives a SIFT feature extraction method based on PCA. Section 4 presents the whole procedure for proposed image registration method. Sections 5 and 6 give experiments and discussion respectively. Finally, the paper is concluded in Section 7.

2 Salient region detection

2.1 Image salient region extraction

Hou proposes a residual spectrum saliency model from the perspective of information theory [22]. He believes that human visual processing mechanism can be interpreted with efficient encoding. From the point of view of information theory, efficient coding decomposes images into two parts. One is the new information, and the other is priori information. The prior information belongs to the redundant information and should be suppressed in the encoding of the system. For the image scene, these redundant information represent the statistical invariant attribute in the environment. By means of removing the remaining components from the image, the detection algorithm of residual spectrum significant component highlights the new information in the image, that is, the image of the significant components. The main idea is to see the image as multi-object projection accumulated under the uniform background and represents it with the image of the two-dimensional Fourier transform.

The amplitude spectrum is usually affected by the characteristics of object regardless of its position, so the original object image is decomposed into weighted sum of complex fundamental waves. If regarding these fundamental waves as a set of features, the amplitude spectrum shows the weight of these features in the image. Using the residual spectrum algorithm is to suppress the larger proportion of features to a larger extent and to enhance the smaller proportion of features.

Firstly, the two-dimensional Fourier transform of image $I(x, y)$ is calculated to obtain amplitude spectrum $A(f)$ and phase spectrum $P(f)$ respectively. It can be described as

$$A(f) = \|F(I(x, y))\| \quad (1)$$

$$P(f) = \frac{F(I(x, y))}{A(f)} \quad (2)$$

where F represents a two-dimensional Fourier operator. The amplitude spectrum $A(f)$ is taken as logarithm, namely

$$L(f) = \log(A(f)). \quad (3)$$

$L(f)$ is put through low-pass filter in frequency domain and get

$$G(f) = h(f) * L(f) \quad (4)$$

where $h(f)$ is the low-pass filter.

$G(f)$ can be seen as a priori information for image. In order to obtain the novel information in image, the total information can be subtracted from the priori information, that is, the residual spectrum can be expressed as

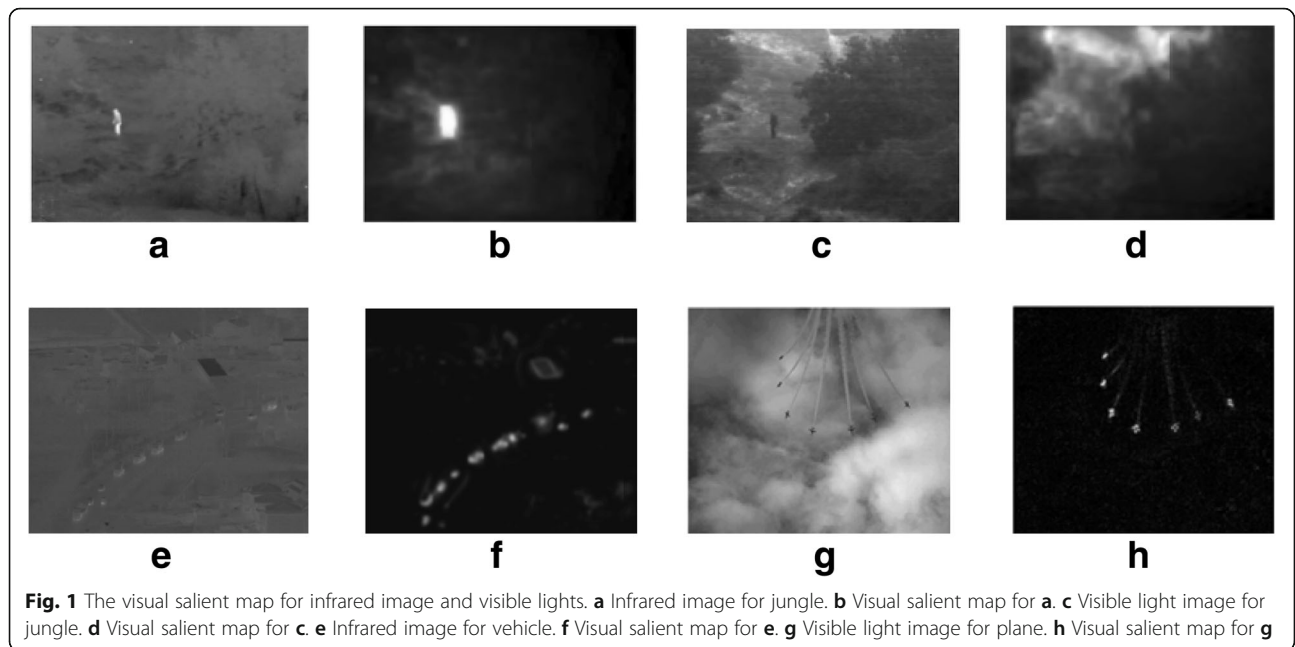
$$R(f) = L(f) - G(f) = L(f) - h(f) * L(f). \quad (5)$$

The amplitude spectrum of image shows the size of each sinusoidal component after Fourier transform, and the phase spectrum denotes the position of these components. The recovery image in view of phase signal corresponds to the dramatic change for edge and the irregular texture regions in image, and human visual system is just interested in these areas. Visual salient map calculation is as follows:

$$S(x, y) = G(x, y) * F^{-1}[\exp(R(f) + i \cdot P(f))]^2 \quad (6)$$

where $G(x, y)$ is the low-pass filter in spatial domain, $P(f)$ is the phase spectrum obtained by Fourier transform, F^{-1} is the Fourier inverse transform, and $S(x, y)$ is the salient mapping obtained.

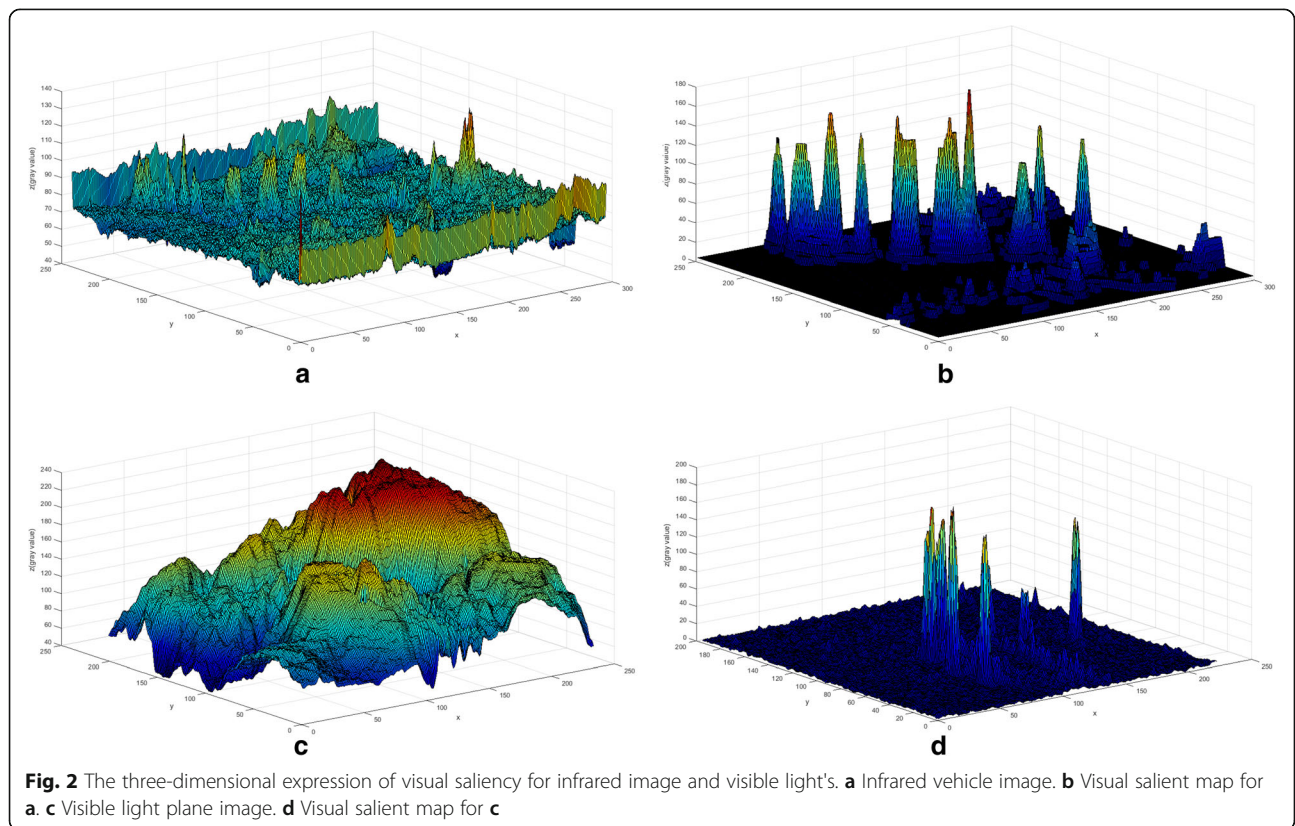
In this paper, the amplitude spectrum experiences normalizing process. This attention selection mechanism can accurately find salient area in image. The extracting result of visual salient map based on amplitude modulation Fourier transform is shown in Fig. 1. The three-dimensional expression of salient map for infrared



vehicle and visible light plane is shown in Fig. 2. It shows that the information in which the human eye is interested can be observed with saliency map. Here, the local extreme points in view of saliency degree in saliency map are called as saliency points.

2.2 Detection of saliency region center based on entropy priority strategy

Entropy is firstly proposed by Rudolf Clausius and applied to the thermodynamics. Then, Shannon introduces the concept of entropy to information theory. In



information theory, the information entropy of vector v is defined as

$$E(v) = - \sum_{i=1}^n p_i \log_2(p_i). \tag{7}$$

In Formula (7), vector v is the set $\{x_1, x_2, \dots, x_n\}$, and the probability of $x_i \in v$ is $p_i = p(x_i)$. The concept of information entropy can be introduced into digital image processing. Taking the range of gray value [0,255] as an example, the gray levels from 0 to 255 are considered as 256 random events. Pixels with different gray levels appear randomly in the image, and the probabilities of occurrence are independent of each other. The image entropy is expressed as bit mean value of the image gray level set and denotes the average information of the image source. Therefore, the image entropy can be used as the scale of the local region information. According to the salient map that extracted before, N salient points can be extracted by taking advantage of the salient threshold. However, these points founded by this method are usually focused on the strongest salient region, so the effect of image understanding is not ideal. In order to compress search space for image registration, the detection for salient region center based on entropy strategy is adopted.

The idea of entropy priority strategy is to take the local entropy of salient point as the scene representative measurement. The strategy firstly selects salient points on the basis of salient map to decrease feature's search space. However, the number of these salient points is tremendous and the position is too centralized; therefore, it is necessary to make the further screening to enhance salient point's effectiveness and information abundance. The following steps are needed to adopt entropy priority strategy to screen the salient point and roughly locate the central position of salient region:

Step 1: Extract local extreme point set in salient map on the basis of salient degree:

$$S = \{(x, y) | (x, y) = \arg \max(G(x, y))\} \\ (x, y) \in \Omega. \tag{8}$$

In Formula (8), Ω is local region as the center of (x, y) , and $G(x, y)$ is the gray value of point (x, y) in salient map for image.

Step 2: Compute image entropy $E(x, y)$ for the points in set S on Ω and rearrange the points on the basis of $E(x, y)$ and get a new sequential set:

$$Q(x_1, y_1) = \arg \max(E(x, y)) \\ Q(x_i, y_i) = \arg \max_{(x, y) \notin R} (E(x, y)) \tag{9}$$

$$(x, y) \in \Omega, R = \bigcup_{m=1}^{i-1} S_m$$

In Formula (9), $Q(x_1, y_1)$ is the first element in set Q and S_m is the point set which has been selected by entropy value.

Step 3: Pick up the first N points in set Q as the centers of salient regions.

The entropy can be used as the measurement of image's local region information, so it can be assumed that the image entropy in same region remains roughly invariable under the same scale. In order to make the same region's scale remain similar between different images, the maximum value strategy of entropy is used to scale matching and find out the region size corresponding to scale:

$$\Omega = \arg \max(E(\Omega_i)). \tag{10}$$

In Formula (10), Ω_i is the pixel set whose range is from 3×3 to 10×10 with a step size of 1. Experimental results of localization of salient region center based on entropy priority strategy are shown as Fig. 3. The white cross line represents the region center point. It can be seen that the central points of these regions can cover most of salient regions of the image. In this experiment, the size of Ω is 7×7 and N is 15.

3 SIFT feature extraction based on principal component analysis

Based on the entropy priority strategy, the center points of image salient regions are extracted, and then, the accurate location of salient image regions is achieved through these center points. The method is mainly divided into four steps:

Step 1: The morphological dilation operator is used to these discrete center points, and the corresponding image region is obtained. Some regions which are connective will be merged. The dilation operator of grayscale morphology is expressed as

$$(f \oplus b)(x, y) = \max\{f(x-x', y-y') + b(x', y') | (x', y') \in D_b\} \tag{11}$$

where f is grayscale image, b is structure element, and D_b is definition domain for b .

Step 2: The minimum outer rectangle for dilation area is obtained as the salient region of image, and then, the value of entropy for the region is calculated.

Step 3: The salient regions of two registered images are cut according to the entropy value, and the unmatched regions are removed.

Step 4: SIFT features are extracted from salient regions of images.

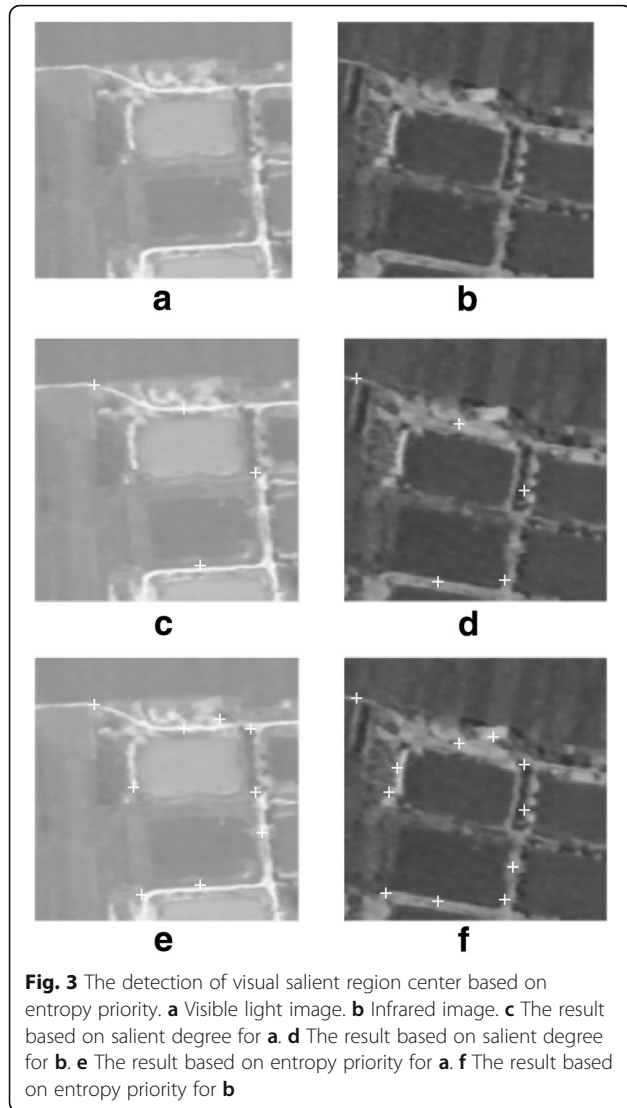


Fig. 3 The detection of visual salient region center based on entropy priority. **a** Visible light image. **b** Infrared image. **c** The result based on salient degree for **a**. **d** The result based on salient degree for **b**. **e** The result based on entropy priority for **a**. **f** The result based on entropy priority for **b**

The SIFT algorithm carries out the extreme value detection in both scale space and two-dimensional plane space and locates these key positions accurately. Then, the main direction of these points is calculated according to the gradient direction of the neighborhood points for the key point position to realize the invariance of geometric transformation and rotations. Generally, three steps are required for generating SIFT feature descriptor:

- Step 1: Establish scale space.
- Step 2: Locate feature points and calculate the main direction.
- Step 3: Represent SIFT feature descriptor with vector.

The vector description dimension of each SIFT feature is 128. The SIFT feature descriptor is a local

feature of image which maintains invariance to rotation transformation, scale zoom, and brightness change. It remains stable to viewpoint change, affine transformation and noise to some extent. However, the size of the SIFT feature descriptor is too large, so the extraction of SIFT feature is time consuming. Besides, there exists redundant among each dimension information for SIFT feature, and the correctness of registration is decreased if the redundancy is not eliminated.

Principal component analysis (PCA) is a statistical method to construct a few major indexes from multiple indexes. PCA can reduce the dimensionality of SIFT feature descriptor vector and improve the matching efficiency.

Let $X_k = [x_1, x_2, \dots, x_n]^T$ be the k sample vector, and n is the vector dimension. The N sample vectors consist of matrix X whose covariance matrix is $R(X)$. The main steps of principal component extraction are as follows:

Step 1: The eigenvalues and eigenvectors for R are calculated, and the eigenvalues are ranked in descending order of $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$, and its corresponding eigenvectors are denoted as $\beta_1, \beta_2, \dots, \beta_n$.

Step 2: Determine the first m principal components y_1, y_2, \dots, y_m as follows:

$$\sum_{j=1}^m \lambda_j / \sum_{i=1}^n \lambda_i \geq q$$

$$y_i = B_i^T X, \quad i = 1, 2, \dots, m. \quad (12)$$

In Formula (12), B is the principal eigenvector matrix composed of the first m eigenvectors β_i . The number of principal components m and the first m principal component vector y_i is determined by the q value. M is usually selectable so that the range of q is around (85%, 95%).

The principal component vector of R generates the same space as the original eigenvector. Any vector in the original eigenvector space can be represented as a linear combination of the principal component vectors. Through principal component analysis, the correlation between the components of the original vector X can be eliminated and the features with less information can be removed. The method of using PCA which decreases the dimensionality of the traditional SIFT feature descriptor is as follows:

- Step 1: Calculate the 128 dimensional SIFT feature descriptor of the key points for all salient regions of the image to be registered and make the feature descriptors whose quantity is n as the sample to write the sample matrix $[x_1, x_2, \dots, x_n]^T$.

Step 2: Calculate the average eigenvector of n samples

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (13)$$

Step 3: Calculate the difference between the feature vectors and the average features, and obtain the difference vectors:

$$d_i = x_i - \bar{x}, i = 1, 2, \dots, n. \quad (14)$$

Step 4: Construct the covariance matrix

$$C = \frac{1}{n} \sum_{i=1}^n d_i d_i^T = \frac{1}{n} Q Q^T \quad (15)$$

where Q is $[d_1, d_2, \dots, d_n]$.

Step 5: Calculate eigenvalues λ_i and eigenvector e_i for the covariance matrix whose number are all 128.

Step 6: According to the order from large to small, arrange 128 eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{128}$ and the corresponding eigenvectors $[e_1, e_2, \dots, e_{128}]$.

Step 7: Choose the feature vectors corresponding to the first maximal t eigenvalues as the principal components.

Step 8: Construct a $128 \times t$ matrix A whose columns consist of eigenvectors.

Step 9: Project the original 128 dimensional SIFT descriptor based on Formula (16) into the n dimension subspace, that is, obtain the descriptor of PCA-SIFT

$$y_i = x_i \cdot B. \quad (16)$$

In Formula (16), the size of matrix B is $128 \times t$ and the size of x_i is 1×128 , so the calculated result of Formula (16) is the matrix of $1 \times t$. Each y_i is a feature description of t dimension, and the original 128 dimensional SIFT feature description is decreased to t dimension PCA-SIFTs.

4 The proposed image registration method

Image registration framework which combines visual salient model and SIFT can be divided into the following sections:

Step 1: Use amplitude modulated Fourier transform to construct visual salient map, then obtain salient information while ignoring the background information which is not significant.

Step 2: Calculate the information entropy of local extreme point's neighborhood in salient map to rearrange these points, then take the front extreme points which are rearranged as the center of the salient region.

Step 3: Achieve the position of image's salient region with morphological dilation operator and connective region combination.

Step 4: Select the key matching points for salient region with SIFT algorithm and eliminate the redundancy of the SIFT feature vectors with PCA, then get the compressed SIFT feature of image registration.

Step 5: Use random sample consensus method [24] to remove the error matching.

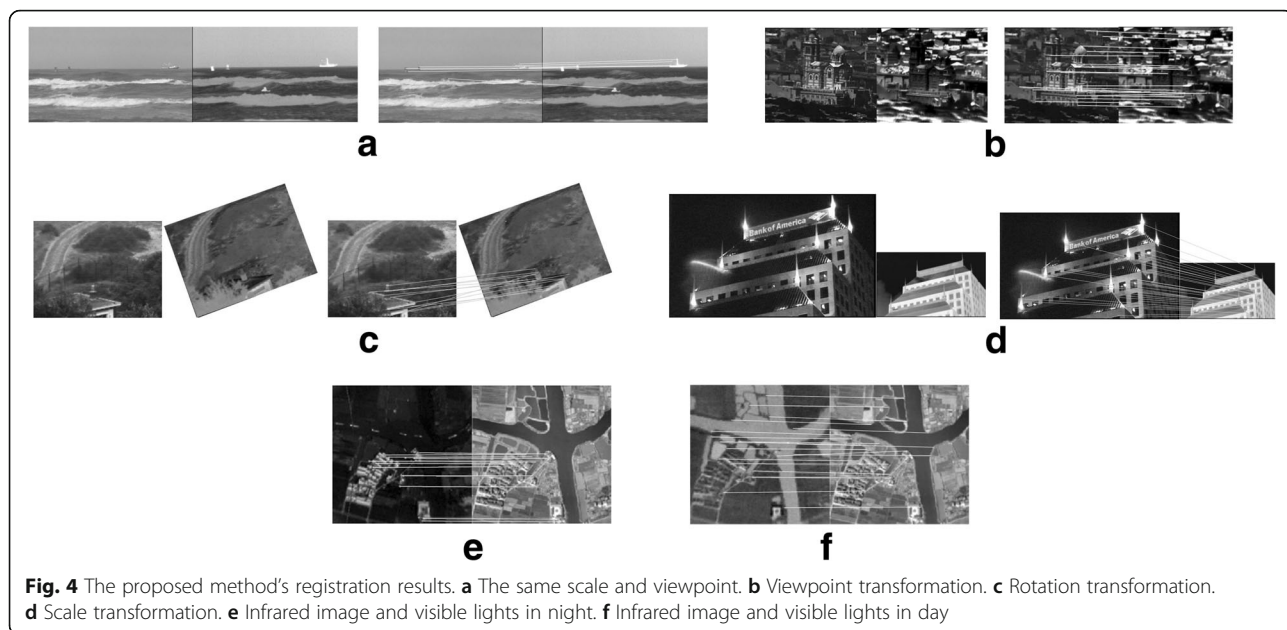
Step 6: Calculate the affine transformation model parameters and apply it to image registration process. The affine transformation model is expressed as

$$Y = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} = AX + t \quad (17)$$

where $X = [x, y]$ is the coordinate for point in registered image and Y is the coordinate for corresponding point in referenced image. $A = [a, b; c, d]$ is the nonsingular matrix, and $t = [t_x, t_y]$ is the coordinate offsets for direction x and y . Therefore, the affine transformation model has six freedoms.

5 Experiments

The experimental hardware platform is a desktop computer, CPU main frequency is 3.4GHz, memory is 4 GB, and software platform is matlab 2014b. Due to having no universal image database for registration field between infrared image and visible lights, this paper makes experiment with computer simulation image and aerial video to verify the validity of the proposed algorithm. The representative images which are transformed by perspective, rotation, scale, and illumination are used as test. Then, the presented algorithm is applied to the procedure of registration between infrared and visible light image in two actual videos. The detection of salient region center based on entropy priority strategy is obtained according to Formulas (8), (9), and (10). The value for parameter N is 15. Parameter t is selected as 20 in SIFT feature extraction based on principal component analysis, so the size of matrix B is 128×20 and the size of x_i is 1×128 . The calculated result for y_i according to Formula (16) is a 20-dimensional feature descriptor; therefore, the original 128-dimensional SIFT feature descriptor is decreased to 20 dimension.



6 Results and discussion

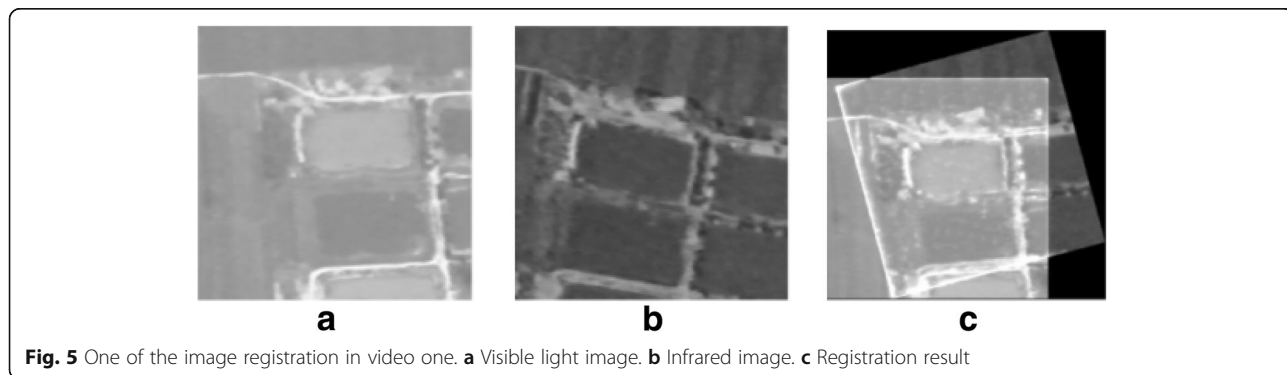
6.1 Results

Figure 4 shows six groups of experimental results. Two groups of infrared and visible light video image sequence which have 1000 frames respectively are engaged in registration experiment. Figures 5 and 6 show registration results with the algorithm in this paper. Table 1 denotes the comparison of registration accuracy and efficiency between PCA-SIFT algorithm and SIFTs based on visual saliency. Table 2 and Figs. 7 and 8 compare the performance of SIFT, GLOH (Gradient Location Orientation Histogram) [25], Harris corner [26], and this proposed method in infrared and visible light image registration.

6.2 Discussion

Figure 4a shows the registration results for infrared and visible image with the proposed method under the condition of the same scale and viewpoint. The images in Fig. 4b have certain viewpoint transformation. Due to the feature point positioning of salient

region center based on entropy priority strategy, the information of SIFT feature descriptor is dense in neighborhood center region for feature point; however, the information description is sparse in salient region center's peripheral region, so the proposed algorithm is robust to the viewpoint transform in some way. In Fig. 4c, there is 20° of rotation between registration images. The structural feature descriptors possess with rotation invariance and most feature points can achieve the correct matching because of adding the main direction when constructing descriptors and selecting feature point's neighborhood according to the main direction. Figure 4d shows the experimental results for scale invariance. It can be seen that the proposed algorithm has some robust to scale transformation. Figure 4e, f respectively shows the registration results for infrared and visible light image during day and night. It can be seen that this algorithm also has certain robustness for illumination changes.



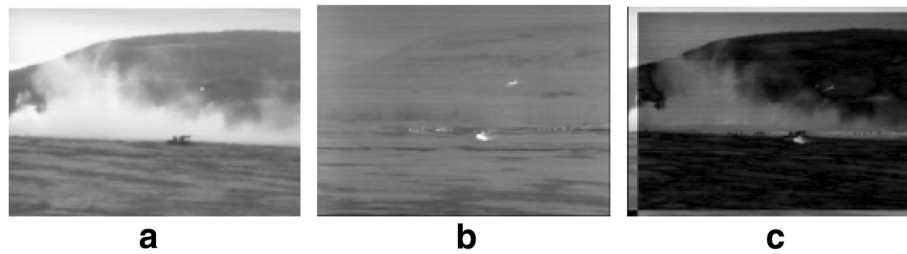


Fig. 6 One of the image registration in video two. **a** Visible light image. **b** Infrared image. **c** Registration result

Figures 5 and 6 show registration results towards the actual videos with the proposed algorithm. Although the contrast exists a big difference between infrared image and visible light image, the proposed algorithm is not sensitive to it and can obtain better registration results. Due to the rough consistency of salient region between video frames, the matching results are mainly concentrated in the shared salient area of two registration images. Moreover, the entropy priority strategy solves the shortcomings that visual salient points are over-concentrated in a certain extent, therefore, salient regions extracted have been represented independently.

The comparison of registration accuracy and efficiency between PCA-SIFT algorithm and SIFTs based on visual

saliency is shown in Table 1. The following conclusions can be drawn:

- (1) The PCA-SIFT algorithm based on visual saliency has very stable registration performance in accuracy rate whether the image is transformed with viewpoint, rotation, scaling, or illumination.

Table 1 The registration result between PCA-SIFT algorithm and SIFTs based on visual saliency

	Registration algorithm	Matching pairs	Accurate pairs	Accuracy rate(%)	Time (s)
Fig. 4a	SIFT based on visual saliency	13	11	84.62	0.537
	The proposed algorithm	9	8	88.89	0.101
Fig. 4b	SIFT based on visual saliency	28	25	89.29	1.138
	The proposed algorithm	21	20	95.24	0.172
Fig. 4c	SIFT based on visual saliency	18	15	83.33	0.837
	The proposed algorithm	12	10	83.33	0.112
Fig. 4d	SIFT based on visual saliency	31	26	83.87	1.316
	The proposed algorithm	26	23	88.46	0.187
Fig. 4e	SIFT based on visual saliency	16	12	75	0.726
	The proposed algorithm	10	8	80	0.105
Fig. 4f	SIFT based on visual saliency	23	18	78.26	1.028
	The proposed algorithm	18	16	88.89	0.167

Table 2 The registration result comparison with some algorithms

	Registration algorithm	Matching pairs	Accurate pairs	Accuracy rate (%)	Time (s)
Fig. 4a	SIFT	18	13	72.22	0.715
	GLOH	19	15	78.95	0.652
	Harris	33	22	66.67	0.578
	The proposed algorithm	9	8	88.89	0.101
Fig. 4b	SIFT	32	27	84.38	1.312
	GLOH	30	25	83.33	1.208
	Harris	48	33	68.75	1.157
	The proposed algorithm	21	20	95.24	0.172
Fig. 4c	SIFT	24	19	79.17	1.138
	GLOH	26	21	80.77	1.087
	Harris	43	29	67.44	0.958
	The proposed algorithm	12	10	83.33	0.112
Fig. 4d	SIFT	38	31	81.58	1.498
	GLOH	41	34	82.93	1.511
	Harris	62	44	70.97	1.482
	The proposed algorithm	26	23	88.46	0.187
Fig. 4e	SIFT	21	15	71.43	0.907
	GLOH	19	15	78.95	0.843
	Harris	30	19	63.33	0.829
	The proposed algorithm	10	8	80	0.105
Fig. 4f	SIFT	27	20	74.07	1.226
	GLOH	28	22	78.57	1.138
	Harris	48	34	70.83	1.082
	The proposed algorithm	18	16	88.89	0.167

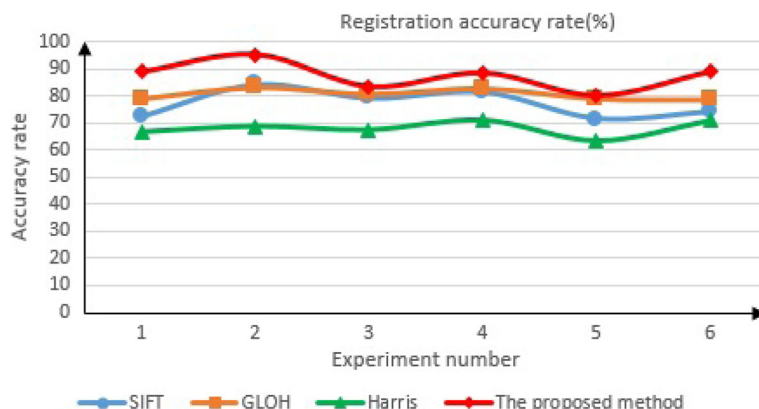


Fig. 7 The curve of registration accuracy comparison with some algorithms

(2) The PCA-SIFT algorithm based on visual saliency is the least time in registration efficiency.

Table 2 and Figs. 7 and 8 compare the performance of SIFT, GLOH, Harris corner, and this proposed method in infrared and visible light image registration. It can be seen that the method in this paper is slightly better in accuracy rate and is obviously lower than other methods in registration time. The proposed method considers not only global information but also local information for image. The method eliminates redundancy information through the visual saliency map extraction effectively and reduces time consuming generally while salient scene region is prominent. The salient region extraction based on the entropy priority strategy covers the significant information in image better. The extraction of local features in image which combines time domain with frequency has better fault tolerance and invariance to translation, direction, scale, and so on. Experimental results show that the use of these features can realize registration effectively between infrared and visible light image.

7 Conclusions

The fast registration method between infrared and visible light image which fuses the visual saliency and SIFT feature can segment the representative image area accurately. Furthermore, the proposed method inherits SIFT feature’s good invariance to viewpoint, rotation, illumination, translation, and scale transformation. On this basis, PCA method reduces the dimension of SIFT feature vector descriptor furtherly and ensures real-time performance for image registration. The simulation image and actual video prove that the proposed method can realize robust registration effectively between infrared and visible light image. Compared with other classical algorithms, this paper’s method has higher registration accuracy rate and faster registration speed. Based on visual saliency region extraction, the strategy of multi-feature fusion will improve accuracy and stability for image registration furtherly, which will be the focus of the following research.

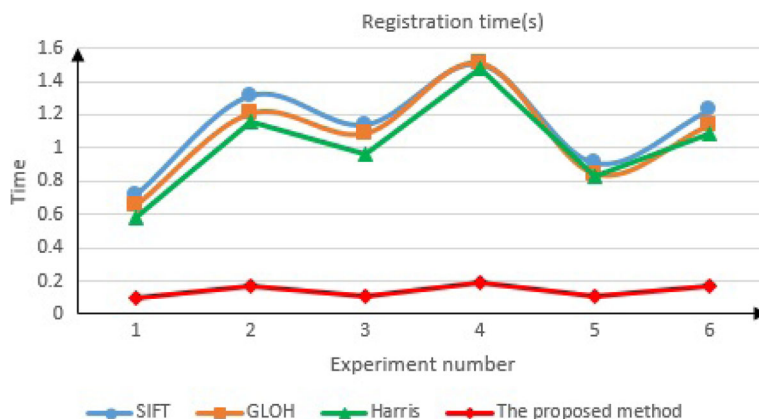


Fig. 8 The curve of registration time comparison with some algorithms

Abbreviations

GLOH: Gradient location orientation histogram; PCA: Principal component analysis; SIFT: Scale invariant feature transform

Acknowledgements

The authors thank the editor and anonymous reviewers for their helpful comments and valuable suggestions.

Funding

This work was supported in part by a grant from the National Natural Science Foundation of China (No. U1504610), Henan Natural Science Foundation (No. 162300410095), and Henan Science and Technology Project (No.172102210039, No.172102410071).

Availability of data and materials

The data used and/or analysed during the current study are available from the corresponding author on reasonable request.

About the authors

Gang Liu was born in Luoyang, Henan, P.R. China, in 1974. He received the Ph.D. degree from Northwestern Polytechnical University, P.R. China. Now, he works in the College of Information Engineering, Henan University of Science and Technology. His research interests include machine vision and system simulation.

Zhonghua Liu was born in Zhengzhou, Henan, P.R. China, in 1975. He received the Ph.D. degree from Nanjing University of Science and Technology, P.R. China. Now, he works in the College of Information Engineering, Henan University of Science and Technology. His research interests include machine vision and system simulation.

Sen Liu was born in Luoyang, Henan, P.R. China, in 1972. He received the Ph.D. degree from Nanjing University of Science and Technology, P.R. China. Now, he works in the College of Information Engineering, Henan University of Science and Technology. His research interests include machine vision and system simulation.

Jianwei Ma was born in Luoyang, Henan, P.R. China, in 1965. He received the Ph.D. degree from Nanjing University of Science and Technology, P.R. China. Now, he works in the College of Information Engineering, Henan University of Science and Technology. His research interests include machine vision and system simulation.

Fei Wang was born in Luoyang, Henan, P.R. China, in 1976. He received the Ph.D. degree from Northwestern Polytechnical University, P.R. China. Now, he works in the College of Information Engineering, Henan University of Science and Technology. His research interests include machine vision and system simulation.

Authors' contributions

All authors take part in the discussion of the work described in this paper. The author GL wrote the first version of the paper. The author ZL did part experiments of the paper; SL, JM, and FW revised the paper in different versions of the paper, respectively. The contributions of the proposed work are as follows: To our best knowledge, our work is the first one to apply the scale invariant feature transform based on visual saliency in registration of infrared and visible images. The method adopts amplitude modulation Fourier transform to construct saliency map. Then, the image salient points as centers for salient regions are achieved preliminarily by salient threshold and reordered based on the strategy of entropy priority. Aiming at the abstracted salient region for image, PCA (principal component analysis)-SIFT algorithm is proposed, which can produce the compressed SIFT registration features and largely reduce the computational cost of image registration. The experimental results indicate that the method has good invariance under image scale, rotation, translation, and illumination variation and can realize effective registration between infrared image and visible light's. All authors read and approved the final manuscript.

Ethics approval and consent to participate

The Academic Board of Information Engineering College, Henan University of Science and Technology.

Competing interests

The authors declare that they have no competing interests. And all authors have seen the manuscript and approved to submit to your journal. We

confirm that the content of the manuscript has not been published or submitted for publication elsewhere.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 27 March 2018 Accepted: 23 May 2018

Published online: 08 June 2018

References

1. J Kern, M Pattichis, Robust multispectral image registration using mutual information models. *IEEE Trans. Geosci. Remote Sens.* **45**(5), 1494–1505 (2007)
2. Z Youwen, G Kun, M Xianghu, et al., Infrared and visual image registration based on mutual information with a combined particle swarm optimization-Powell search algorithm. *OPTIK* **127**(1), 188–191 (2016)
3. JH Lee, YS Kim, D Lee, et al., Robust CCD and IR image registration using gradient-based statistical information. *IEEE Signal Processing Letters* **17**(4), 347–350 (2010)
4. J Klimaszewski, M Kondej, M Kawecki, et al. Registration of infrared and visible images based on edge extraction and phase correlation approaches [C]. 4th International Conference on Image Processing and Communications, 153–162, 2013.
5. H Li, Z Aiwu, H Shaoxing, A registration scheme for multispectral system using phase correlation and scale invariant feature matching. *J Sens* **2016**, (2016). <https://doi.org/10.1155/2016/3789570>
6. R SEA, V Sanchez, G Prince, et al., Registration of thermal and visible light images of diseased plants using silhouette extraction in the wavelet domain. *Pattern Recogn.* **48**(7), 2119–2128 (2015)
7. GA Bilodeau, P St-Onge, R Garnier, in *Computer Vision and Pattern Recognition Workshops, 2011 IEEE Computer Society Conference on. Silhouette-based features for visible-infrared registration* (2011), pp. 20–25
8. H Du, Z Liu, H Song, L Mei, Z Xu, Improving RGBD saliency detection using progressive region classification and saliency fusion. *IEEE Access* **4**, 8987–8994 (2016)
9. D Firmenichy, M Brown, S Susstrunk, in *18th IEEE International Conference on Image Processing. Multispectral interest points for RGB-NIR image registration* (2011), pp. 181–184
10. A Sima, B Aleksandra, et al., Optimizing SIFT for matching of short wave infrared and visible wavelength images. *Remote Sens.* **5**(5), 2037–2056 (2013)
11. P Wang, Z-g Qu, P Wang, et al., A coarse-to-fine matching algorithm for FLIR and optical satellite image registration. *IEEE Geosci. Remote Sens. Lett.* **9**(4), 599–603 (2012)
12. D Xiaojie, L Erqi, Y Jie, et al., Visible and infrared automatic image registration based on SLER. *Journal of Infrared and Millimeter Waves* **33**(1), 90–97 (2014)
13. F Xiaolei, W Wei, L Zhi, et al., Weighed-Hausdorff distance using gradient orientation information for visible and infrared image registration. *OPTIK* **126**(23), 3823–3829 (2015)
14. R Istenic, D Heric, S Ribaric, in *The 14th International Conference on System, Signal and Image Processing, and 6th EURASIP Conference Focused on Speech and Image Processing, Multimedia Communications and Services. Thermal and visual image registration in Hough parameter space* (2007), pp. 106–109
15. S Verstockt, C Poppe, V Hoecke, et al., Silhouette-based multi-sensors smoke detection coverage analysis of moving object silhouettes in thermal and visible registered images. *Mach. Vis. Appl.* **26**(6), 1243–1262 (2012)
16. J Han, EJ Pauwels, P de Zeeuw, Visible and infrared image registration in man-made environments employing hybrid visual features. *Pattern Recogn. Lett.*, 42–51 (2013)
17. L Zhigang, L Yan, P Qua, et al., A robust multi-level scene matching algorithm for infrared and visible light image. *Acta aeronautica et astronautica sinica* **31**(6), 1185–1195 (2010)
18. L Xiang, Z Zunshang, S Yang, et al., Multimodal image registration based on orientation-moment. *Journal of national university of defense technology* **37**(1), 153–158 (2015)
19. L Gang, W Guangyu, Z Heng, et al., Multiresolution scene matching algorithm for infrared and visible images based on non-subsampled contourlet transform. *Journal of System Simulation* **37**(1), 1795–1804 (2016)

20. DG Lowe, Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **60**(2), 91–110 (2004)
21. C Siagian, L Itti, Rapid biologically-inspired scene classification using features shared with visual attention. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(2), 300–312 (2007)
22. Hou Xiaodi, Zhang Liqing, Saliency detection: a spectral residual approach. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ, USA: IEEE, 2438–2445, 2007.
23. Chen D Y, Han P, Wu C D. Frequency spectrum modification: a new model for visual saliency detection[C]. *Proceedings of the 7th International Symposium on Neural Networks*. Berlin, Springer Verlag, 90–96, 2010.
24. D Nannan, L Yanying, Z Ye, et al., Fast image registration based on SURF-DAISY algorithm and randomized kd trees. *Journal of Optoelectronics. Laser* **23**(7), 1395–1402 (2012)
25. K Mikolajczyk, C Schmid, Scale and affine invariant interest point detectors. *Int. J. Comput. Vis.* **60**(1), 63–86 (2004)
26. K Mikolajczyk, C Schmid, Indexing based on scale invariant interestpoints[C]. *Proceedings of the 8th international conference on computer Vision*. Vancouver, Canada. IEEE, 525–531 (2001)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
