An Improved Laparoscopic Image Registration Algorithm Based on Sift

Jiujiang Zhou, Jianxu Mao, and Xiaoyan He

College of Electrical and Information Engineering, Hunan University, Changsha, China
jiujiang.z@gmail.com, mao_jianxu@126.com

Abstract. Image registration is a recognized difficulty and many people are working on it to make their algorithms more efficient and robust. In image-guided surgical and interventional procedures, the registration precision and real time effect are both quite important for the following accurate tissue deformation recovery and 3D anatomical registration as well as navigation. This article uses the radon-transform and bidirectional matching approach on SIFT(Scale Invariant Feature Transform) which is aiming at the registration in laparoscopic binocular vision. Finally, we test the new algorithm and give better experiment results by comparing with other common methods.

Keywords: Medical Image Registration, Radon Transform, Bidirectional Matching, SIFT.

1 Introduction

Medical image registration is the basis to realize medical image information fusion and three-dimensional reconstruction and so on, it has been widely applied in disease diagnosis and preoperative planning, etc. Because its better adaptability to position change, gray level change, image distortion and complex space transform, the registration based on feature point is the current mainstream direction as well as development trend. A good feature detector or feature tracking technique is important for accurate tissue deformation recovery, 3D anatomical registration and navigation in computer assisted MIS(Minimally Invasive Surgical) procedures. Although there are a variety of methods which have been developed for image registration, there is no higher-performance and shorter time-consuming algorithm available for fitting great changes like scale, rotation, affine and projection.

Reference [1] expounds the review of medical image registration. Reference [2] raises Harris operator, which has a good performance with rotation and illumination changes but it is also variational with scale changes. Reference [3] gives the Scale Invariant Feature Transform(SIFT) algorithm and it is one of the best methods for feature detection and matching due to the invariance of scale, rotation, illumination, geometric distortion and resolution differences. Reference [4] introduces an Speeded Up Robust Features(SURF) which is based on integral image haar derivation and reduces the operation time. Reference [5] comes up with PCA-SIFT to reduce the dimension. However, this algorithm is not fully affine invariant and the projection matrix requires a series of representative

S. Li et al. (Eds.): CCPR 2014, Part II, CCIS 484, pp. 179–188, 2014.

images. Recently, Harris and Hessian corner point detectors have also been extended to detect affine-invariant regions in [6] and [9], respectively. In [7], the salient region detector is proposed, where the local maximum in affine transformation space is detected by measuring the entropy of pixel intensity histograms computed for elliptical regions. Reference [8] uses a watershed like segmentation algorithm to detect Maximally Stable Extremal Regions (MSER) which is closed under the affine transformation of image coordinates and invariant to affine transformation of intensity. A comparison of these affine region detectors can be found in [9].

This article introduces SIFT in detail and make some improvements based on it. Section1 describes the image registration application on MIS and the commonly used image registration method based on feature points. Section2 gives the main process of SIFT algorithm. Section3 and Section4 demonstrate our modified ideas on two aspects: dimension reduction of the descriptor based on radon transform; a novel bidirectional matching approach. Section5 displays the experimental results and analysis, the method is applied on laparoscopic binocular vision images and the results show greater accuracy and faster matching. Section6 gives the conclusions and the deficiency of our method as well as the direction of further improvements. The purpose of this paper is to reduce the operating time as well as increase the efficiency of the algorithm for image registration in MIS and our proposed idea is tested on `in vivo` video sequences from robotic assisted MIS procedures.

2 Scale Invariant Feature Transform

The scale invariant feature transform (SIFT) algorithm, developed by Lowe[3,10,11], is an algorithm for image features matching which is invariant to image translation, scaling, rotation and partially invariant to illumination changes and affine projection. The main process of the algorithm is composed of the following four parts:

2.1 Scale-Space Local Extrema Detection

Lowe uses gaussian difference function to identify the key points which are scale and orientation invariant. The scale space of the image is defined as the function $L(x, y, \sigma)$, which is convolved by the variable scale gaussian function

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y), \tag{1}$$

in which the gaussian function is:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x-\frac{m}{2})^2 + (y-\frac{n}{2})^2}{2\sigma^2}} \tag{2}$$

Differential gaussian scale space *D(x,y,σ)* is defined as the convolution between the original image and the adjacent differential scale gaussian function which includes a constant factor κ. These local extremum of gaussian difference image are regard as the feature points on the corresponding scale spatial domain, and the function *D(x,y,σ)* can be expressed by:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) \times I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \tag{3}$$

2.2 Accurate Positioning of Feature Points

By calculating the fitting function

$$D(X) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X \tag{4}$$

and further catching the extreme point

$$\hat{X} = -\left(\frac{\partial^2 D}{\partial X^2}\right)^{-1} \frac{\partial D}{\partial X}, \tag{5}$$

we can get the corresponding extreme value

$$D(\hat{X}) = D + \frac{1}{2} \frac{\partial D^T}{\partial X}, \tag{6}$$

thus obtain the local optimal point by revising X and get rid of the weak feature points in which

$$\left| D(\hat{X}) \right| < 0.03. \tag{7}$$

Meanwhile, accurate location and scale of the candidate feature points are acquired. The edge points also need to be removed by Hessian matrix. Assuming that Tr(H), Det(H) are adding result and product of the eigenvalues, γ is the ratio of these matrix eigenvalues, for the sake of testing if the main curvature is less than a certain threshold γ, just check if the equation

$$\frac{Tr(H)^2}{Det(H)} < \frac{(\gamma + 1)^2}{\gamma} \tag{8}$$

is right, which γ is defined as 6~10 .

2.3 The Generation of SIFT Feature Descriptor

Collect the gradient information and direction distribution property in the 3 σ neighborhood of the gaussian pyramid image, in which the radius is defined as:

$$radius = \frac{3\sigma_{oct} * \sqrt{2}(d+1) + 1}{2}, \tag{9}$$

and confirm the gradient direction in the known keypoints. Ascertain the maximum value of histogram which is divided into 36 bins is the main direction of the keypoint and only keep the bins which have the size of the 80% of the peak value in the main direction or greater than it as the assistant direction of this keypoint which is to keep the descriptor's invariance of scale. The value and direction of the gradient at keypoint (x,y) are described as

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \tag{10}$$

$$\theta(x,y)=\arctan\frac{L(x,y+1)-L(x,y-1)}{L(x+1,y)-L(x-1,y)} \quad . \tag{11}$$

At last, the 128 dimensional vector is normalized for the invariance of the influence of illumination change, and after the processing the vector can be described as

$$L=(l_1,l_2,...,l_{128})\,, \tag{12}$$

in which

$$l_i=\frac{h_i}{\sqrt{\sum_{j=1}^{128}h_j}} \quad . \tag{13}$$

3 Descriptor Dimension Reduction

In this part, a novel formed descriptor is put forward based on radon transform [12]. The intrinsic quality of the registration based on keypoints between two images is that just matching two image subregions centered on every feature keypoint. As a consequence, first acquire the scale, orientation and position information of the feature points by using gaussian difference scale space. Second, applying the radon transform on the image to be processed in a series of straight line and catch the new descriptors for matching. By appropriately selecting some specific angles which are used for calculating radon transform value, the purpose of dimension reduction is achieved. At last, add feature point direction angle into the integral function to reduce the computational cost caused by rotation. The details are as follows.

3.1 Radon Transform Introduction

Assuming that the the main direction of feature points detected by SIFT feature is Y', while the angle is defined as θ between Y' and y coordinate system. In original SIFT algorithm, in order to transform the tiny displacement of the feature points which results in the change of feature variable value, Lowe make a gaussian weighted smoothing method. When making a pixel gradient amplitude statistics, the sampling region near the feature point is highlighted. However, the proportion is dropping with the distance between the sampling region and the center point increasing. This paper will achieve the same purpose by changing product factor included by integral function of the image radon transform, and adopt the product factor $\frac{1}{1+\sqrt{x}}$. The integral function is :

$$R(x)=\int_{S'}\frac{1}{1+\sqrt{x}}I(x,y) \tag{14}$$

in which S' represents the integral region with the keypoint centered and is same to the statistic region in the original SIFT algorithm. The radon transform of the image is shown as Fig1.(a).

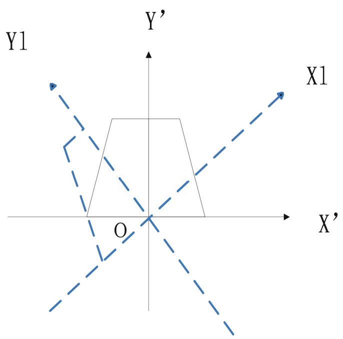

Fig. 1. (a) The radon transform

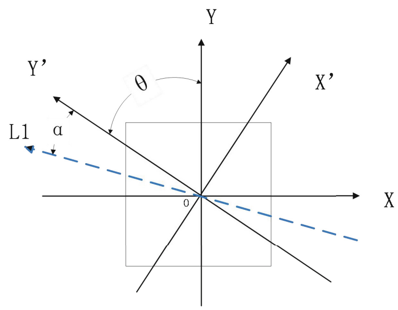

Fig. 1. (b) The main direction of feature points and rotation Angle

3.2 Seek For D-Dimension Characteristic Vector Space

As shown in the Fig. 1(b), for the purpose of achieving d-SIFT feature vector space, feature point's principle orientation Y' is deemed to be as the reference direction. Make other d-1 lines $(l_1, l_2, ..., l_{d-1})$ and thereinto, L_1 is shown in the Fig. 1(b). The included angle between two adjacent lines

$$\alpha = \frac{2\pi}{d}, \tag{15}$$

so the radon transform of main direction line Y' belongs to image $I(x, y)$ can be expressed by:

$$R_\theta(x') = \int_S \frac{1}{1+\sqrt{x'\cos\theta - y'\sin\theta}} I(x'\cos\theta - y'\sin\theta, x'\sin\theta + y'\cos\theta)dy' \cdot \tag{16}$$

Similarly, the radon transform of $I(x, y)$ along the lines $L_1, L_2, ..., L_{d-1}$ can be shown as:

$$R_{\omega_n}(x_{l_n}) = \int_S \frac{1}{1+\sqrt{x_{l_n}\cos\omega_n - y_{l_n}\sin\omega_n}} I(x_{l_n}\cos\omega_n - y_{l_n}\sin\omega_n, x_{l_n}\sin\omega_n + y_{l_n}\cos\omega_n)dy_{l_n} \tag{17}$$

in which

$$\omega_n = \theta + n\alpha, n = 1,2,...,d-1, \tag{18}$$

Fig. 1(b) shows the main direction of feature point of the image $I(x, y)$ and radon transform of line L_1.

3.3 Selecting the Appropriate Integral Angle

In order to apply the idea into our method, choosing the integral angle α as $\frac{\pi}{15}$, and normalized the length of eigenvector to unit form. Then, achieve 30 dimensional radon-SIFT descriptor:

$$R_{\theta}(x^{'}), R_{\omega_1}(x_{l_1}), R_{\omega_2}(x_{l_2}),..., R_{\omega_{29}}(x_{l_{29}}) \tag{19}$$

4 Bidirectional Matching Approach

4.1 Define the Distance of the Eigenvectors

Generally speaking, the best candidate match for a keypoint in I_1 is found by measuring the nearest distance between keypoints. The minimum euclidean distance is defined for the invariant descriptor vector and used to match points form I_1 to I_2. Supposing that the feature descriptor generated respectively from two images are:

$$R_{\theta}(x_1^{'}), R_{\omega_1}(x_{1l_1}), R_{\omega_2}(x_{1l_2}),..., R_{\omega_{29}}(x_{1l_{29}});\ R_{\theta}(x_2^{'}), R_{\omega_1}(x_{2l_1}), R_{\omega_2}(x_{2l_2}),..., R_{\omega_{29}}(x_{2l_{29}}), \tag{20}$$

thus euclidean distance will be:

$$d = \sqrt{\sum_{i=0}^{29}(R_{\omega_i}(x_{1l_i}) - R_{\omega_i}(x_{2l_i}))^2}\ . \tag{21}$$

This paper still choose the Best Bin First(BBF) method to search the nearest and next nearest neighbor feature points , and the ratio between them is set a threshold value T. If the ratio is lower than T, the matching is conformed successfully.

4.2 Bidirectional Match

Suppose that F and G are two set of descriptors generated by I_1 and I_2 separately, and the size of them are M and N. Considering the feature similarity, We define the measurement function by $S = \Omega(F, G)$ where S is the similarity measurement matrix, in which Ω is method of similarity measurement. After obtaining S matrix, select a arbitrary element of F and make the comparison with all the elements in G, and then find out the pair which S obtain the extremum. Next, traverse in turn all the other elements of F as well as obtaining the correspondences, in which this process is a one-way mapping from F to G. The correspondence can be roughly shown as:

$$correspondings(S_{\max(1\to N)}) = [\mathrm{maximum}(F \to G)_{1\to \mathrm{N}}, index(F \to G)_{1\to N}], \tag{22}$$

on the contrary, one-way mapping from G to F:

$$correspond\,ings(S_{\max(1\to M)}) = [\mathrm{maximum}(G \to F)_{1\to \mathrm{M}}, index(G \to F)_{1\to M}], \tag{23}$$

where the maximum $(F \to G)_{1\to N}$ represents the max value of S, index $(F \to G)_{1\to N}$ represents the corresponding descriptors, and vice versa. Finally, gain the best K matches which is approximately shown as figure2 form the two descriptor sets. The K should also meet the Eq.(24)

$$[index(F \to G)_{1\to K}]\&[index(G \to K)_{1\to K}] == K \tag{24}$$

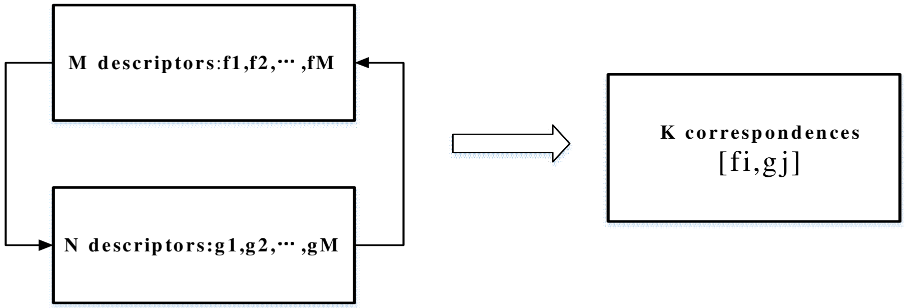

Fig. 2. Two-way matching diagram

4.3 A Mixed Similarity Measurement Strategy

The previous section discusses the matching method and the way of similarity measurement Ω is used. In the field of medical image registration, the evaluation of registration results does not exist the so-called absolute gold standard, which is due to illumination change, imaging equipment parameters, etc. With the continuous development of research on image registration there have been some good similarity measure functions, such as quadratic sum of the gray differentials, mutual information, image correlation and so on. Due to the application for registration in the laparoscopic binocular vision images are mostly center symmetrical, the global measurement function Ω is determined by combining image gray level information and feature descriptors at the same time. Therefor, we also set up a weight value β for optimum proportion. The measurement function is:

$$\Omega[I_1, I_2] = \beta\Omega_\beta(p(x), q(y)) + (1-\beta)\Omega_{1-\beta}(f(x), g(y)). \tag{25}$$

Among these symbols,

$$\begin{cases} \Omega_\beta(p(x), q(y)) = \dfrac{1-|p(x)-q(y)|}{R} \\ \Omega_{1-\beta}(f(x), g(y)) = \dfrac{< f(x) \bullet g(y) >}{|f(x)||g(y)|} \end{cases}, \tag{26}$$

$p(x)$ and $q(y)$ are the gray information of image I_1 and I_2, $f(x)$ and $g(y)$ are feature descriptors of these two images respectively, R is the gray statistics region.

$<\bullet>$ is the distance between the descriptors as described above, β is adjustment coefficient. $\Omega_{\beta}(p(x),q(y))$ is gray similarity measure function, $\Omega_{1-\beta}(f(x),g(y))$ is descriptor similarity measure function. To continuously verify registration results parameter β needs to be adjusted. In this paper we determine that the value is 0.25 based on the experimental results of specific laparoscopic environment.

5 Experimental Results

The sensitivity of a given registration algorithm is estimated as defined sensitivity as the ratio of correct matchings and all the detected matchings.

$$sensitivity = \frac{\#correct\_matches}{\#correspondences} \quad , \tag{27}$$

In case of illumination, scale and rotation changes, test the new improved algorithm on the given data [13], as is shown in Fig3 (a), which is aiming at the improvement of characteristic dimension and matching strategy. At last, the behavior of new proposed algorithm is compared with several other classical algorithms. Experimental platform: CPU: 2.7 GHz core i5, memory 6G, Windows 7, Matlab2010. Through trial and error, the datum obtained are calculated as follows:

Table 1. The comparison diagram of different algorithm

	SIFT		SURF		PCA-SIFT		MSER		OUR METHOD	
	Time (s)	Sensitivity (%)	Time (s)	Sensitivity (%)	Time (s)	Sensitivity (%)	Time (s)	Sensitivity (%)	Time (s)	Sensitivity (%)
Illumination change	4.81	91.5	3.15	83.1	2.67	87.2	3.62	68.8	2.15	92.8
Scale change	3.38	84.6	1.98	87.7	2.19	82.8	2.97	76.6	1.86	89.7
Orientation change	6.07	73.1	3.12	65.8	3.8	66.9	4.5	84.6	3.25	81.5

The Table1 shows that under the three common transformations, compared with the original sift algorithm and other commonly used algorithm, the new 30-d radon bidirectional-SIFT has better overall promotion at the accuracy of the matching as well as time reduced by about 30%. The second way SURF depends on local pixel gradient direction much, and it may also be invalid as long as the layers of the pyramid are not closely enough. Although PCA-SIFT reduces the running time, but the projection matrix still needs some typical images. For MSER, its main advantage is good affine invariance. At the same time, this algorithm needs more improvement on instantaneity. In general the new algorithm improves the efficiency especially on registration accuracy under conditions of keeping good scale invariance and the original sift matching ratio. The results of the experiment effect using our novel algorithm can be visible as the following figures:

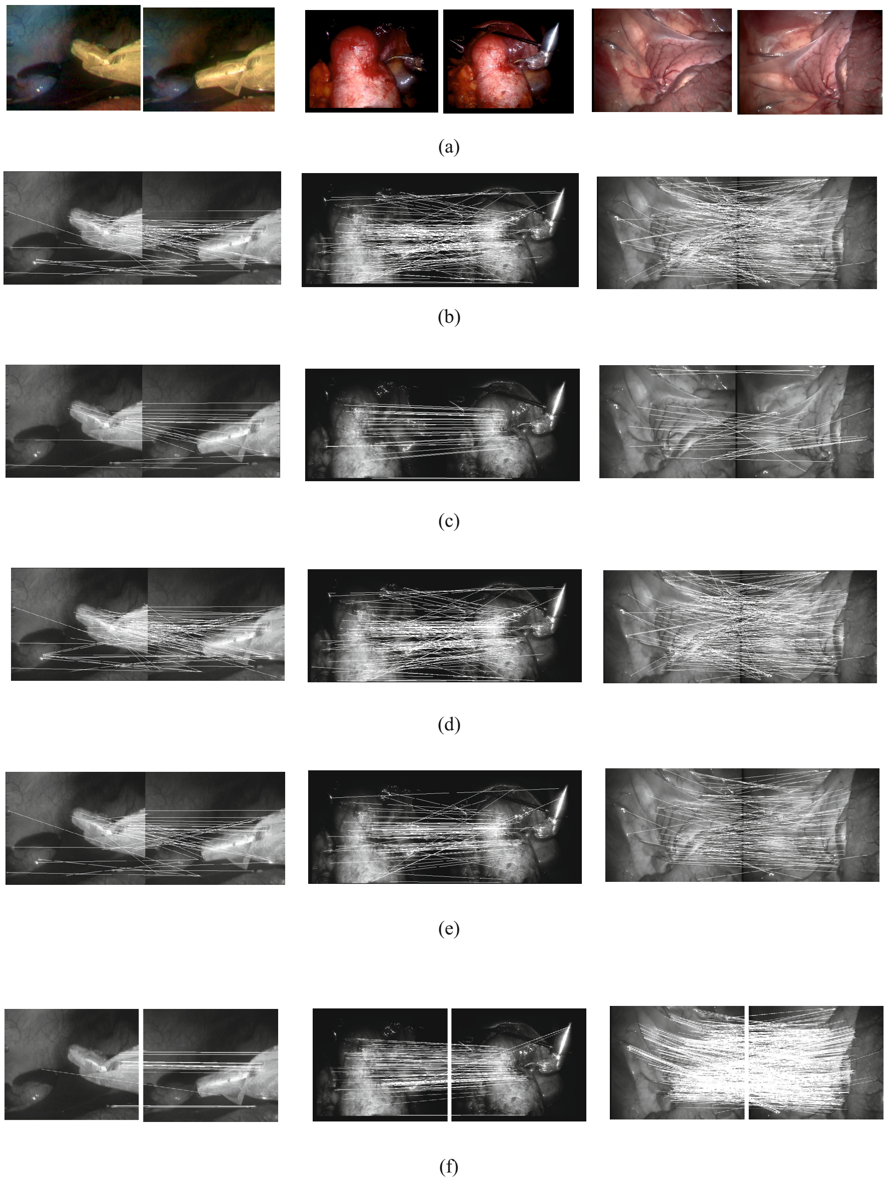

Fig. 3. Test images and results. (a) Original Images. (b) SIFT Results. (c) SURF Results. (d) PCA-SIFT Results. (e) MSER Results. (f) Our Improved Algorithm Results.

6 Conclusions and Discussion

In this paper, a novel feature detection and matching algorithm based on SIFT is presented. Experimental results have shown that the proposed improvement of the original method is capable of making a more accurate matching and taking less time. This work improves the SIFT contraposing the matching accuracy as well as operating time and provides some new thinking for improving the efficiency of algorithm. The SIFT itself maybe also have limitations, maybe it relies too much on the gradient direction of local pixels when striving for main direction which results in matching errors increasing; just make up the scale error when building dimension by interpolation. In the following work, the emphasis will continue to be put on searching for more efficiency algorithm and prepared for the higher precision registration requirement in 3d reconstruction. Of course it can also be used for *in vivo* tracking, significant deformation and intraoperative registration.

Acknowledgements. This work is supported by National Natural Science Foundation of China (61072121, 60835004, 61271382), Hunan Provincial Natural Science Foundation of China (12JJ2035) and the Fundamental Research Funds for the Central Universities, Hunan university.

References

[1] Maintz, J.B., Viergever, M.A.: A survey of medical image registration. Medical Image Analysis 2(1), 1–36 (1998)
[2] Harris, C., Stephens, M.: A combined corner and edge detector. In: Alvey Vision Conference, vol. 15, p. 50 (1988)
[3] Lowe, D.G.: Distinctive Image Features from Scale-invariant Keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
[4] Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
[5] Ke, Y., Sukthankar, R.: PCA-SIFT: A more distinctive representation for local image descriptors. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004, vol. 2, pp. II-506–II-513. IEEE (2004)
[6] Mikolajczyk, K., Schmid, C.: Scale & affine invariant interest point detectors. International Journal of Computer Vision 60(1), 63–86 (2004)
[7] Kadir, T., Zisserman, A., Brady, M.: An affine invariant salient region detector. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3021, pp. 228–241. Springer, Heidelberg (2004)
[8] Matas, J., Chum, O., Urban, M., et al.: Robust wide-baseline stereo from maximally stable extremal regions. Image and Vision Computing 22(10), 761–767 (2004)
[9] Mikolajczyk, K., Tuytelaars, T., Schmid, C., et al.: A comparison of affine region detectors. International Journal of Computer Vision 65(1-2), 43–72 (2005)
[10] Lowe, D.G.: Object recognition from local scale-invariant features. In: The Proceedings of the Seventh IEEE International Conference on Computer Vision, vol. 2, pp. 1150–1157. IEEE (1999)
[11] Lowe, D.G.: Local feature view clustering for 3D object recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, pp. 682–688 (December 2001)
[12] Kadyrov, A., Petrou, M.: The trace transform and its applications. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(8), 811–828 (2001)
[13] Giannarou, S., Visentini-Scarzanella, M., Yang, G.-Z.: Probabilistic Tracking of Affine-Invariant Anisotropic Regions. IEEE Transactions on Pattern Analysis and Machine Intelligence 99 (2012)