A Fast Specific Object Recognition Algorithm in a Cluttered Scene

Lei Liang[1,5], Jeng-Shyang Pan[2,3,4], Yongjun Zhuang[5]

[1]Department of Computer Science and Technology, Harbin Institute of Technology, China
[2]College of Computer Science and Engineering, Shandong University of Science and Technology, China
[3]Fujian Provincial Key Lab of Big Data Mining and Applications, Fujian University of Technology, China
[4]Department of Information Management, Chaoyang University of Technology, Taiwan
[5]Sanbot Innovation Research Institute, Sanbot Innovation Intelligence Co., LTD, China
lianglei8568@163.com, jspan@cc.kuas.edu.tw, greyzhuang@139.com

Abstract

Specific object recognition technology is an important research component of computer vision and image processing technology and is also used in Industrial Internet and Internet of Things. In recent years, due to the widespread use of visual surveillance systems, specific object recognition technology has been gradually applied in monitoring systems based on image processing. By satisfying image feature invariance under changes in such factors as scale, illumination, and rotation, point-matching methods based on local invariant feature transform (LIFT) have recently become an attractive field for specific object recognition. In this paper, we propose a fast specific object recognition algorithm in a cluttered scene based on LIFT. First, the pyramid level of the reference image that is closest to the scale of the object in the scene image is determined and is referred to as the corresponding level. Next, the resolution of the scene image is increased by enlarging the interpolation based on the corresponding level. Finally, LIFT matches the reference image and interpolated scene image on a single level, reducing false match pairs in addition to considering the number of keypoint restriction problems. The experimental results demonstrate that the proposed algorithm is not only fast but also highly robust compared to existing algorithms. Thus, it can recognize objects correctly.

Keywords: Object recognition, Local invariant features, Feature matching

1 Introduction

Specific object recognition refers to the recognition of a specific object instance (e.g. my cup), instead of a generic object class/category (e.g. a cup), which is usually called object categorization or generic object recognition. At present, the state-of-the-art technology for object recognition is the use of LIFT (locally invariant features transform) of images [1], and the state-of-the-art technology for object classification is deep learning [2]. In this article, we only focus on the object recognition algorithm.

Automatic target detection and recognition is an important area of computer vision. With the development of target recognition technology, many successful applications have been employed in Industrial Internet and have greatly improved its intelligent level [3-4]. The automatic installation of semiconductor devices based on image recognition was employed in the semiconductor industry starting in 1973 [5-6], as shown in Figure 1, the principle of the transistor assembly machine is to identify the object to be assembled by using the LIFT of the image. This application is one of the most successful applications of computer vision technology in industry [7].

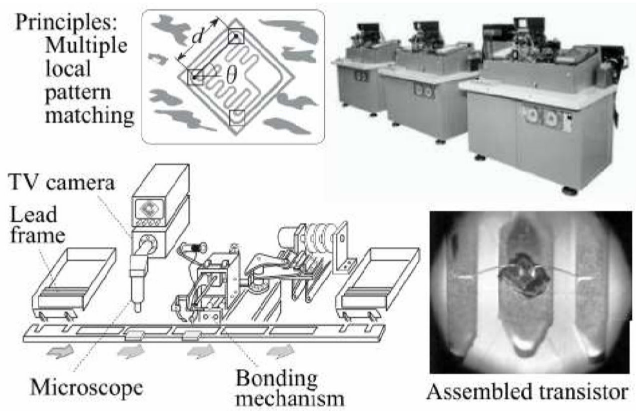


Figure 1. Transistor assembly machine [5]

In addition to the Industrial Internet field, target recognition has a wide range of potential applications in social life. An earlier successful application is a character pattern recognition system associated with office automation [8], which greatly enhanced the efficiency of traditional literature digitization. This

*Corresponding Author: Jeng-Shyang Pan; E-mail: jspan@cc.kuas.edu.tw
DOI: 10.3966/160792642019122007001

technology was subsequently employed for license plate recognition [9-10] in intelligent transportation, currency identification [5] in ATM machines, postal code identification for postal sorting, and fingerprint [11-13] and iris [14] recognition, which is widely used in criminal detection. All of these applications are the part of Internet of Things and result of the development of a target recognition algorithm.

This paper investigates fast specific object recognition technology in a cluttered scene. Specific objects generally have discernible texture, color or shape characteristics, but in cluttered scenes, the features will be subject to varying degrees of damage, such as the influence of light and the occlusion between objects. This paper proposes a specific object recognition algorithm in a cluttered scene for these types of problems. The proposed algorithm is based on the image LIFT model.

The remainder of this paper is organized as follows: Section 2 introduces the LIFT of the image. Section 3 describes the proposed algorithm in detail. Section 4 evaluates the proposed algorithm in cluttered scenes and compares the proposed algorithm with the object recognition algorithm that is also based on the LIFT model. Section 5 summarizes the paper.

2 LIFT of Image

At present, an increasing number of objects play a significant role in human life, such as laptops, mobile phones, U disks and other types of mobile information storage devices. These devices continue to be smaller and more portable, but they are also more likely to be lost and are easy to steal and difficult to control. At the same time, video streaming places a higher real-time performance requirement on the algorithm. LIFT-based object recognition method is currently the state-of-the-art specific object recognition method [1]. The algorithm flow is as follows: First, the reference image containing the object of interest (Figure 2(a)) and the scene image (Figure 2(b)) are collected. Next, the LIFT extraction algorithm is used to detect feature points and extract feature descriptors from both images. Then, the features are matched using their descriptors (Figure 2(c)). Finally, the geometric transformation is estimated, and the outliers are eliminated. This transformation allows us to localize the object in the scene (Figure 2(d)).

2.1 Background

The LIFT method refers to the detection and description of the features that remain unchanged for the various changes in the image, such as geometric transformation, photometric transformation, convolution transformation, and perspective changes. The basic concept of the LIFT method is to extract the essential image attributes. It is independent and can adapt to the specific image content (that is, the feature extracts the adaptive change to describe the same image content when the expression changes). The LIFT method typically has a local support neighborhood. Unlike the classical image segmentation algorithm, the local support neighborhood may be any subset of the image, and the boundaries of the support region do not necessarily correspond to changes in the image's appearance (color or texture). LIFT extraction is generally divided into LIFT detection (typically a point or region) and LIFT description (the expression of the feature region).

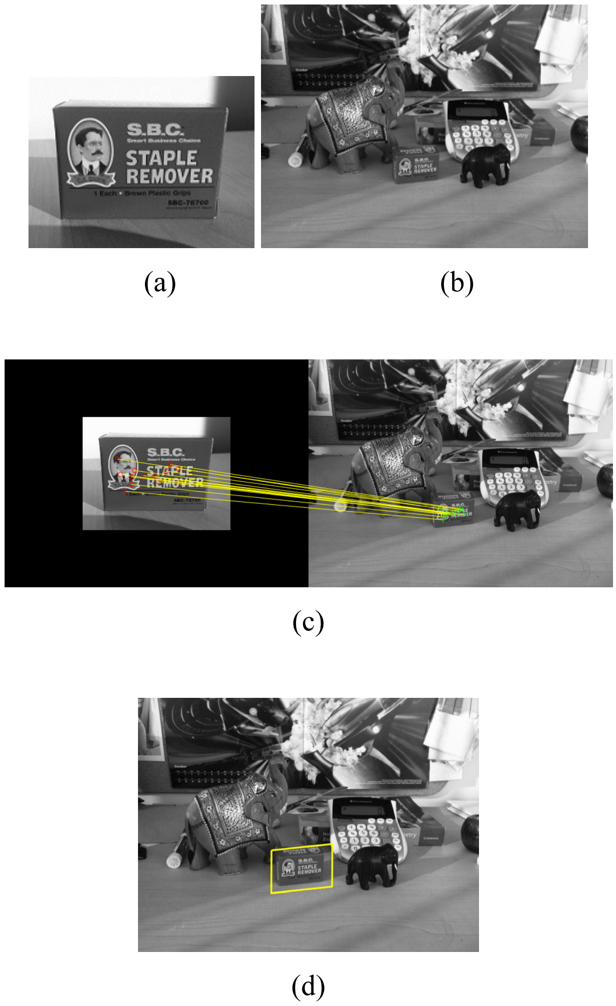


Figure 2. A specific object is recognized by the proposed in a cluttered scene, given a reference image of the object

Feature detection is a prerequisite for feature description. The purpose of feature detection is to locate the point of interest [15-20], blob [21-23], or area [24] in the image. The LIFT detection algorithms presented in the literature can be divided into three types based on the different feature levels: corner invariant feature, blob invariant feature and regional invariant feature.

Feature description is a prerequisite for feature matching. The purpose of feature description is to express the feature information by quantifying the attributes of the feature. Numerous LIFT description methods have been presented in the literature and can be divided into distribution-based feature description methods [21, 25, 26], filtering-based feature description methods [27-28], moment-based feature description methods [29-31] and other description methods.

A recent review of the literature [1] summarized the application of the LIFT method in object recognition, concluding that the best performing object recognition systems are built using the Oriented FAST and Rotated Binary Robust Independent Elementary Features (BRIEF) (ORB) [32] and scale-invariant feature transform (SIFT) [21] methods. The ORB method is faster, whereas the SIFT method is more robust under real-world conditions. In this paper, the proposed method is based on ORB for the real-time requirements of the system, but it greatly enhances the robustness of the ORB method. So next we describe the ORB algorithm in detail.

2.2 ORB Method

The ORB method is a feature point extraction method with orientation-invariant and noise-insensitive properties. It uses the features from accelerated segment test (FAST) [33] to detect the feature points, and an orientation-invariant feature description is proposed based on the improved BRIEF [34]. The ORB method includes three steps, as detailed below.

2.1.1 Corner Detection

In the ORB method, FAST-9 is used as the corner detection method. To maintain the scale invariance, the ORB method introduces an image pyramid when corner detection is performed and then detects the FAST corner point on each pyramid image.

2.1.2 Determine the Orientation of the Corner

In the ORB method, the direction of the corner point is defined as the direction from the center point of the corner point O to the center point of the image block C. The moment of an image block is defined as

$$m_{pq} = \sum_{x,y} x^p y^q I(x,y) \tag{1}$$

The center point of the image block C is defined as

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \tag{2}$$

Then the orientation of the corner $\overrightarrow{OC}$ is

$$\theta = a\tan 2(m_{01}, m_{10}) \tag{3}$$

where *atan*2 is the quadrant-aware version of *arctan*.

2.1.3 Generate the Descriptor of the Feature Point

First, the BRIEF is calculated. In the ORB method, the size of the image block for calculating the BRIEF descriptor is taken as 31×31, and the dimension of the feature vector is taken as 256.

The second is the calculation of BRIEF, according to the pixel (x_i, y_i) selected for calculating the descriptors, $2\times N$ matrix is defined as

$$S = \begin{pmatrix} x_1, \ldots x_n \\ y_1, \ldots y_n \end{pmatrix} \tag{4}$$

The corresponding rotation matrix R_θ is constructed using the orientation of the image block θ, The transformation of S can be obtained S_θ:

$$S_\theta = R_\theta S \tag{5}$$

The steered BRIEF descriptor can be obtained according to the following definition:

$$g_n(p,\theta) := f_n(p) \,|\, (x_i, y_i) \in S_\theta \tag{6}$$

Use S_θ generated feature descriptor to ensure that the orientation of descriptors invariance.

Finally, the steered *BRIEF* is optimized to *rBRIEF*. Because the high variance and low correlation make the feature easier to distinguish, the *ORB* has developed a learning method to restore the variance that is lost when calculating the *BRIEF* descriptor and to reduce the correlation within the binary test subset. The method is as follows:

(1) Calculate the binary test of the entire training image block.

(2) The vector T is formed by sorting all the tests according to the distance

(3) Search by greedy method:

(*a*) Move the first test from T into the result vector R.

(*b*) Get the next test from T, compare it to the test in R, discard it if the correlation is greater than a threshold, and otherwise add it to R.

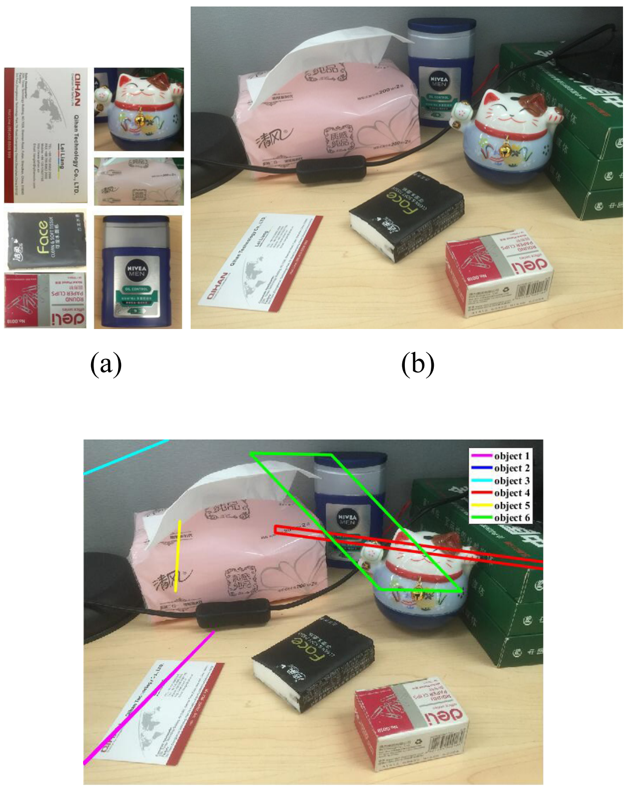
(*c*) Repeat the previous step until there are 256 tests in R. If less than 256, then increase the threshold and then recalculate.

After the above operation, the steered *BRIEF* descriptor is optimized as a *rBRIEF* descriptor and becomes the final feature vector.

3 The Proposed Method

Although we can cite a large number of successful application examples for the target recognition algorithm, this does not mean that this algorithm has become mature. On the contrary, it is not yet a unified and theoretical framework, and there are many challenges in theory and technology. From a theoretical perspective, the theoretical framework of the target recognition algorithm is still in an exploratory phase. Because of the complexity of the human vision system, it is impossible to develop an accurate simulation, and the academic community has much to learn concerning visual theory. In the application field, the ability of perform in real-time and robustness are two primary problems of the target recognition system, particularly in achieving robust and fast target recognition in a complex scene.

Figure 3 shows the results of specific object recognition using the ORB LIFT extraction algorithm. The accuracy of the location is determined by the number of correct matches, i.e., true positives (TPs). As shown in Figure 3c, the locations of all objects are not accurate. The first reason for this inaccuracy is the problem of scale space. Many false match pairs will occur in the cross-scale matching because the image pyramid is used to simulate the scale changes in the original ORB algorithm and this method is not a continuous variation of scale space. The second issue concerns the image resolution. The resolution of the image pyramid decreases layer-by-layer, and thus, it has a significant impact on the number of keypoints that are detected by neighborhood-based corner detection. Moreover, the ORB keypoints of the image pair are restricted to the same number considering the efficiency, but object recognition in a scene is different from other applications (such as wide-baseline matching). A specific object is only a portion of the scene, and thus, the keypoints of the scene image must be more than that of the reference image.



(a) (b)

(c)

Figure 3. ORB matching results

Figure 4 is the proposed algorithm framework, as shown in figure, the essential capability of our method is to determine the corresponding pyramid level of the reference image that is closest to the scale of the object in the scene image. Based on the corresponding pyramid level, we not only avoid cross-scale false matches but also significantly increase the correct number of matches by enlarging the interpolation. The detailed description is as follows:

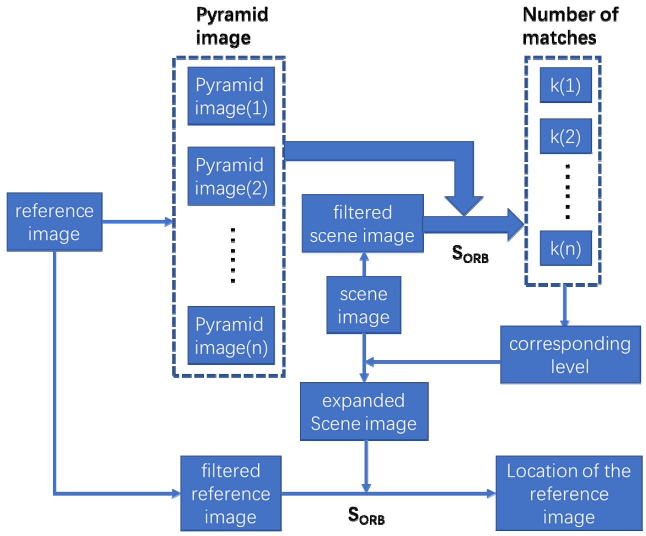


Figure 4. Proposed algorithm framework

Step 1. Establish multi-scale characterization models of reference images. Due to the real-time requirements of the algorithm, create an pyramid image $P_r(x, y, n)$ of the reference image $I_r(x, y)$.

$$P_r(x,y,n) = F_{s(n)}[I_r(x,y)], \qquad n = 0,1,2...N \tag{7}$$

where $F_{s(n)}$ is bilinear interpolation which the scale factor $s(n) = 1/s_{init}{}^{n}$. N is the total number of levels of the pyramid.

Step 2. Apply the ORB matching between $P_r(x, y, n)$ and the scene image $I_s(x, y)$, filtered by $G(x, y, \sigma)$, and obtain the number of matches $k(n)$. A Gaussian filter can reduce the interference caused by other portions of the scene.

$$S_{ORB\,(N_r,\,N_s)}\begin{cases} P_r(x,y,n) \\ I_s(x,y) * G(x,y,\sigma) \end{cases} \Rightarrow k(n), \qquad n = 0,1,2...N \tag{8}$$

where S_{ORB} represents single level matching of ORB. N_r and N_s are restricted keypoint numbers of the reference image and scene image, respectively. Because we want to obtain the corresponding level, we set N_s based on the ratio of the image resolution.

$$N_s = \frac{R_s}{R_r} \cdot N_r \tag{9}$$

where R_r and R_s are the resolution of the reference and scene images, respectively.

Step 3. Assign different weights to *k(n)* based on the scale of the pyramid image. Then, obtain the corresponding level c by taking the maximum value of *K(n)*.

$$K(n) = k(n) \cdot s_{init}^{n}, \quad n = 0,1,2...N \tag{10}$$

$$c = max[K(n)], \quad n = 0,1,2...N \tag{11}$$

Step 4. Calculate the scale factor of the bilinear interpolation according to the corresponding level *c* and obtain the expanded scene image $E_s(x,y)$ by interpolation *F*.

$$E_s(x, y) = F_{s_{init}^{c}}[I_s(x, y)] \tag{12}$$

Step 5. Perform single-level ORB matching between $E_s(x, y)$ and $I_r(x, y)$, filtered by $G(x, y, \sigma)$. Gaussian filtering is used to simulate the loss of details due to expanded interpolation.

$$S_{ORB(N_r', N_s')}\begin{cases} E_s(x, y) \\ I_r(x, y) * G(x, y, \sigma) \end{cases} \tag{13}$$

In addition to the number of restricted keypoints related to the image resolution, there are those related to the content of the image. If the object has less texture, it is easily affected by other objects in the scene that have more texture. Thus, we also must consider image entropy.

$$N_s' = \alpha \cdot \frac{\beta_s}{\beta_r} \cdot \frac{R_s}{R_r} \cdot N_r' \tag{14}$$

where α is a coefficient factor and β_r and β_s are global entropies of the reference and scene images, respectively.

The result obtained using the proposed method for specific object recognition is shown in Figure 5. As shown in the figure, the location of each object in the scene is marked precisely. The corresponding level of each specific object is shown in Table 1. The scene image resolution is 980×735.

Figure 5. The proposed method results

Table 1. Corresponding level of each specific object

Object number	Image resolution	Corresponding level
1	294×482	4
2	343×258	4
3	294×221	3
4	734×695	6
5	487×266	1
6	368×490	4

Although the proposed method uses single-level matching, it can be easily extended to multi-level correspondence matching, thus further increasing the number of TPs. However, as shown in Table 2, this also increases the execution time. In the next experiment, we show that single-level matching meets the recognition requirements.

Table 2. Multi-level correspondence matching results

level	TP	Time (ms)
1	36	57.7
2	90	98.3
3	146	136.9
4	203	171.6
5	269	194.4
6	328	223.2
7	396	233.1
8	443	250.1

4 Experimental Comparison

BRISK [35] is another common feature extraction method based on corner detection and binary description. The proposed method will be compared with ORB and BRISK. A cluttered scene image is shown in Figure 6. The experimental objects are 20 cards with different textures. The resolution of each card is 368×490. Occlusion between the cards is a serious problem.

The experimental parameters are as follows: $N = 7$, $N_r = N_r' = 500$, $s_{init} = 1.2$, $\sigma = 0.3$, and $\alpha = 2$, and ORB and BRISK use the same number of restricted keypoints and default values for the remaining parameters. The TP matches are obtained by M-estimator Sample Consensus (MSAC) [36] to eliminate outliers. The experimental results are shown in Table 3. The gray box indicates that the correct location of the card is marked (see Figure 7). As shown in Table 3, the improvement effect of the proposed algorithm is highly significant. The TP rate of the ORB method is low, and only 30% of the cards are correctly marked. The TP rate of the BRISK method was better than that of the ORB method. However, no card was recognized correctly due to the insufficient number of matches. The proposed method is superior in terms of both the number of matches and TP rate, where the number of matches increased by 36% and the TP rate increased by 8 times compared to the ORB method.

Figure 6. Reference images (368×490) and cluttered scene image (735×980)

Table 3. Comparison of the number of matches and TP rate

Card number	ORB		BRISK		Proposed	
	Match number	TP rate (%)	Match number	TP rate (%)	Match number	TP rate (%)
1	86	5.81	50	/	91	72.53
2	59	6.78	18	12.5	63	55.55
3	72	5.55	29	13.79	96	92.71
4	82	6.09	23	/	124	91.13
5	84	4.76	27	18.52	23	43.48
6	71	23.94	18	27.78	91	67.03
7	84	4.76	32	18.75	112	86.61
8	73	13.69	30	13.33	152	87.5
9	87	4.59	16	18.75	50	62
10	95	16.84	45	8.89	258	87.59
11	80	5	38	/	31	32.26
12	109	13.76	25	/	168	89.29
13	79	11.39	41	/	140	72.86
14	84	5.95	18	16.66	85	63.77
15	67	5.97	28	10.71	83	74.69
16	90	6.67	29	10.34	103	94.17
17	90	4.44	53	5.66	81	81.48
18	74	8.11	16	18.75	79	78.48
19	73	16.44	34	/	234	97.86
20	62	6.45	11	27.27	119	88.24
average	80.05	8.85	29.05	15.84	**109.15**	**75.96**

Figure 7. Recognition results obtained using the proposed method

In practical applications, such as the use of a mobile robot to recognize special objects, the reference image is acquired in advance, and the reference pyramid image can be pre-established. Table 4 shows the time required to execute the other major steps of the experiment. All experiments were performed using a C++ program on a 3.2 GHz processor. Because a complex scale space and feature descriptor template must be constructed, the BRISK method requires more time for feature extraction. The time required for the proposed method to perform feature extraction is similar to that of the original ORB method, but the proposed method does not need to establish an image pyramid and has expanded image scene resolution. Compared with the ORB and BRISK methods, the proposed method requires additional time to determine the corresponding level. However, MSAC requires less time due to its high accuracy. In summary, the proposed method is faster than BRISK and maintains the advantage of the ORB method's speed.

Table 4. Comparison major steps processing time

Methods		Time (ms)
ORB	Feature Extraction	75.9
	Feature matching	21.4
	MSAC	169.5
BRISK	Feature Extraction	268.3
	Feature matching	17.1
	MSAC	139.2
proposed	Find Corresponding	206.4
	Feature Extraction	70.6
	Feature matching	20.7
	MSAC	5.5

5 Conclusion

In this paper, we proposed a fast specific object recognition algorithm for use in evaluating a scene. The proposed method has been tested in a cluttered scene and has been shown to outperform the BRISK and ORB methods. In contrast to the original multi-level matching method, the proposed algorithm uses an expanded image single match to eliminate a portion of the false matches and increase the number of correct matches. The algorithm proposed in this paper can be applied to content-based image retrieval, robot grasping based on vision and other fields in the future.

References

[1] P. Loncomilla, J. Ruiz-del-Solar, L. Martínez, Object Recognition Using Local Invariant Features for Robotic Applications: A Survey, *Pattern Recognition*, Vol. 60, pp. 499-514, December, 2016.

[2] Y. LeCun, Y. Bengio, G. Hinton, Deep Learning, *Nature*, Vol. 521, pp. 436-444, May, 2015.

[3] J. S. Pan, L. P. Kong, T. W. Sung, P. W. Tsai, V. Snasel, a-Fraction First Strategy for Hirarchical Wireless Sensor Neteorks, *Journal of Internet Technology*, Vol. 19, pp. 1717-1726, June, 2018.

[4] J. S. Pan, C. Y. Lee, A. Sghaier, M. Zeghid, J. Xie, Novel Systolization of Subquadratic Space Complexity Multipliers Based on Toeplitz Matrix-Vector Product Approach, *IEEE Transactions on Very Large Scale Integration Systems*, Vol. 27, pp. 1614-1622, July, 2019.

[5] M. Ejiri, Machine Vision in Early Days: Japan's Pioneering Contributions, *Asian Conference on Computer Vision*, Tokyo, Japan, 2007, pp. 35-53.

[6] S. Kashioka, M. Ejiri, Y. Sakamoto, A Transistor Wire-bonding System Utilizing Multiple Local Pattern Matching Techniques, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-6, No. 8, pp. 562-570, August, 1976.

[7] E. N. Malamas, E. G. M Petrakis, M. Zervakis, L. Petit, J. D. Legat, A Survey on Industrial Vision Systems, Applications and Tools, *Image And Vision Computing*, Vol. 21, No. 2, pp. 171-188, February, 2003.

[8] L. Roberts, Pattern Recognition with an Adaptive Network, *Proceedings of the Institute of Radio Engineers*, Vol. 48, No. 3, pp. 66-70, February, 1960.

[9] K. Takahashi, T. Kitamura, M. Takatoo, Y. Kobayashi, Y. Satoh, Traffic Flow Measuring System by Image Processing, *Iapr Workshop on Machine Vision Applications*, Tokyo, Japan, 1996, pp. 245-248.

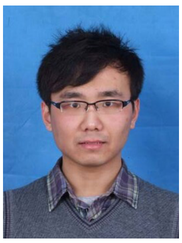
[10] C. N. E. Anagnostopoulos, I. E. Anagnostopoulos, I. D. Psoroulas, V. Loumos, E. Kayafas, License Plate Recognition from Still Images and Video Sequences: A Survey, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 9, No. 3, pp. 377-391, May, 2008.

[11] C. L. Lin, K. C. Fan, Biometric Verification Using Thermal Images of Palm-dorsa Vein Patterns, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 14, No. 2, pp. 199-213, March, 2004.

[12] N. Miura, A. Nagasaka, T. Miyatake, Extraction of Finger-vein patterns Using Maximum Curvature Points in Image

Profiles, *IEICE TRANSACTIONS on Information and Systems*, Vol. 90, No. 8, pp. 1185-1194, August, 2007.
[13] D. Maltoni, D. Maio, A. K. Jain, S. Prabhakar, *Handbook of Fingerprint Recognition*, Springer Science & Business Media, 2009.
[14] K. W. Bowyer, K. Hollingsworth, P. J. Flynn, Image Understanding for Iris Biometrics: A Survey, *Computer Vision and Image Understanding*, Vol. 110, No. 2, pp. 281-307, May, 2008.
[15] H. P. Moravec, *Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover*, Stanford University, 1980.
[16] C. Harris, M. Stephens, A Combined Corner and Edge Detector, *Alvey Vision Conference*, Vol. 3, pp. 147-151, March, 1988.
[17] J. Shi, Good Features to Track, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE*, Seattle, WA, USA, 1994, pp. 593-600.
[18] T. Lindeberg, Feature Detection with Automatic Scale Selection, *International Journal of Computer Vision*, Vol. 30, No. 2, pp. 79-116, November, 1998.
[19] K. Mikolajczyk, C. Schmid, Scale & Affine Invariant Interest Point Detectors, *International Journal of Computer Vision*, Vol. 60, No. 1, pp. 63-86, October, 2004.
[20] S. M. Smith, J. M. Brady, SUSAN: A New Approach to Low Level Image Processing, *International Journal of Computer Vision*, Vol. 23, No. 1, pp. 45-78, May, 1997.
[21] D. G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91-110, November, 2004.
[22] P. R. Beaudet, Rotationally Invariant Image Operators, *Proc. Int. Conf Pattern Recognition*, Tokyo, Japan, 1978, pp. 579-583.
[23] H. Bay, T. Tuytelaars, L. V. Gool, Surf: Speeded Up Robust Features, *European Conference On Computer Vision*, Graz, Austria, 2006, pp. 404-417.
[24] J. Matas, O. Chum, M. Urban, T. Pajdla, Robust Wide-baseline Stereo from Maximally Stable Extremal Regions, *Image and Vision Computing*, Vol. 22, No. 10, pp. 761-767, September, 2004.
[25] Y. Ke, R. Sukthankar, PCA-SIFT: A More Distinctive Representation for Local Image Descriptors, *Computer Vision and Pattern Recognition, IEEE,* Washington, DC, USA, 2004, pp. 506-513.
[26] K. Mikolajczyk, C. Schmid, A Performance Evaluation of Local Descriptors, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 27, No. 10, pp. 1615-1630, October, 2005.
[27] W. T. Freeman, E. H. Adelson, The Design and Use of Steerable Filters, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 13, No. 9, pp. 891-906, September, 1991.
[28] A. Baumberg, Reliable Feature Matching across Widely Separated Views, *Computer Vision and Pattern Recognition*, Hilton Head Island, SC, USA, 2000, pp. 774-781.
[29] M. K. Hu, Visual Pattern Recognition by Moment Invariants, *IRE Transactions on Information Theory*, Vol. 8, No. 2, pp. 179-187, February, 1962.
[30] H. Shu, L. Luo, J. Caatrieux, Moment-based Approaches in Imaging: Basic Features, *IEEE Engineering in Medicine and Biology Magazine*, Vol. 26, No. 5, pp. 70-74, September, 2007.
[31] S. O. Belkasim, M. Shridhar, M. Ahmadi, Pattern Recognition with Moment Invariants: A Comparative Study and New Results, *Pattern Recognition*, Vol. 24, No. 12, pp. 1117-1138, December, 1991.
[32] E. Rublee, V. Rabaud, K. Konolige, G. Bradski, ORB: An Efficient Alternative to SIFT or SURF, *IEEE International Conference on Computer Vision*, Barcelona, Spain, 2011, pp. 2564-2571.
[33] E. Rosten, T. Drummond, Machine Learning for High-Speed Corner Detection, *European Conference on Computer Vision*, Graz, Austria, 2006, pp. 430-443.
[34] M. Calonder, V. Lepetit, C. Strecha, P. Fua, Brief: Binary Robust Independent Elementary Features, *European Conference on Computer Vision*, Heraklion, Crete, Greece, 2010, pp. 778-792.
[35] S. Leutenegger, M. Chli, R. Y. Siegwart, BRISK: Binary Robust Invariant Scalable Keypoints, *IEEE International Conference on Computer Vision*, Barcelona, Spain, 2011, pp. 2548-2555.
[36] P. H. S. Torr, A. Zisserman, MLESAC: A New Robust Estimator with Application to Estimating Image Geometry, *Computer Vision and Image Understanding*, Vol. 78, No. 1, pp. 138-156, April, 2004.

Biographies



Lei Liang is a post-doc of Computer Science and Technology, Shenzhen Graduate School, Harbin Institute of Technology, China. He received a Ph.D. degree in Automation and Information Engineering from Xi'an University of Technology in 2015. His research interests mainly focus on image processing, tone mapping, and local invariant feature.

Jeng-Shyang Pan received the B.S. degree in Electronic Engineering from the National Taiwan Institute of Technology, Taiwan, in 1986, the M.S. degree in Communication Engineering from National Chiao Tung University, Taiwan, ROC in 1988, and the Ph.D. degree in Electrical Engineering from the University of Edinburgh, UK, in 1996. Currently, he is a Professor in the Fujian Provincial Key Lab of Big Data Mining and Applications and the Shenzhen Graduate School, Harbin Institute of Technology, China. His current research interests include pattern recognition, information security and data mining.

Yongjun Zhuang received the B.S. degree in Electronic Engineering from Hunan University of Technology in 2004. Currently, he is the chief technology officer of Sanbot Innovation Intelligence Co., Ltd. His research interests mainly focus on EDA, ISP (Image signal process), High Speed Fiber and Wireless communication, Motion Control and machine vision. He is family with IOT (Internet of Things) system and AI (Artificial Intelligence).