ISSN: 0976-9102(ONLINE)
DOI: 10.21917/ijivp.2014.0126

RESEARCH ON FEATURE POINTS EXTRACTION METHOD FOR BINARY MULTISCALE AND ROTATION INVARIANT LOCAL FEATURE DESCRIPTOR

Hongwei Ying[1], Jiatao Song[1], Jinhe Wang[1], Xuena Qiu[1], Wang Wei[1] and Zhongxiu Yang[1]
School of Electron and Information Engineering, Ningbo University of Technology, China
E-mail: [1]yinghongwei@163.com

Abstract

An extreme point of scale space extraction method for binary multiscale and rotation invariant local feature descriptor is studied in this paper in order to obtain a robust and fast method for local image feature descriptor. Classic local feature description algorithms often select neighborhood information of feature points which are extremes of image scale space, obtained by constructing the image pyramid using certain signal transform method. But build the image pyramid always consumes a large amount of computing and storage resources, is not conducive to the actual applications development. This paper presents a dual multiscale FAST algorithm, it does not need to build the image pyramid, but can extract feature points of scale extreme quickly. Feature points extracted by proposed method have the characteristic of multiscale and rotation Invariant and are fit to construct the local feature descriptor.

Keywords:

Features Extraction, Multiscale, Feature Descriptor, Corner Detection, Rotation Invariant

1. INTRODUCTION

With the development of digital image storage and processing technology, how to make better use of these mass image data becomes a difficult problem to us. The human visual system can recognize thousands classes of objects accurately and quickly, even changing in view angle, illumination, occlusions and background disturb. The capability of human visual system is attempted to artificially simulate by computer vision techniques for core research tast including image information processing, image target recognition which provide important support for the application of intelligent monitoring, video retrieval, motion video analysis, unmanned vehicle, automatic object detection, etc. After decades of development, image target recognition technology is still a challenging problem although has made considerable progress.

Single target in one image always present many different appearances due to the large variation range of view angle and illumination condition. The same class of target may also have many different representation of the subclass. Several targets in one image may be covered arbitrarily each other. Complex background and target's distortion to a certain degree also increases the difficulty of target recognition. Comprehensive effects of these factors' make the complexity of the problem is far beyond people's anticipate, so image target recognition become a scientific problem and always attract researchers.

In order to make the image target recognition method with the adaptability to these factors above-mentioned, local invariant features have been adopted by many researchers for building target model. An image target distributes its specific local invariant features in its many positions, so even if the target is disturbed by background or covered partly by others targets, remainder local invariant features can still characterize a target with strong robustness. As an advantage to target global features, target recognition method based on local invariant features needn't segment the target from image, so in recent years, local invariant feature extraction has become a very active research direction.

Local invariant features must possess repeated extraction characteristics with transform condition, sufficiency in features' information, rich in quantity and high efficiency of feature extraction etc., performance to meet the need of application of image target recognition.

In terms of the classification stage of a full target recognition tast, if the extracted features can be faster, better to describe the target, then the alternatives schemes of target classifier will be more and the design work would be easier, thus it can reduce the complexity of target recognition system to a certain extent and improve its utility.

In general, extracting work for local feature descriptor has the following 3 steps: (1) building the image pyramid, the image is decomposed in different scales; (2) getting the extreme points in image scale space; (3) selecting features in the neighborhood of extreme points to complete feature descriptor. Among which, the first steps need to consume a large amount of computation and storage resources, is not conducive to the development of practical applications.

The method proposed by this paper abandons traditional ideal of the first stage above-mentioned, takes a new ideal for the extraction of scale extreme point.

2. LOCAL FEATURE DESCRIPTOR

Features extraction is the foundation of target recognition tast, image local feature descriptor overcome the shortcoming of global feature descriptor which is susceptible to masking, disturbing and other issues, is a hotspot of current research in the visual field [1,2].

After image information acquisition, the image's original data which essentially is the pixel's color value matrix, which can be transformed to the effective characteristics for specific classification tasks through data processing. Summary many recognition algorithms, commonly used types of features are: (1) Corner feature of an image target; (2) Edge contour feature of an image target; (3) color feature including color information distribution and the mean in an image; (4) Texture features: texture complexity and texture direction of each area in an image; (5) Mathematical characteristics: pixel's or block's correlation and other features which physical meaning is not obvious. These features must be robust at first that means when test samples occur translation, rotation, scaling conditions relative to training samples, the classifier can still identify objects by using these features. Secondly, feature extraction algorithm must take into account the time complexity, if the calculation time is too long, usually it is very difficult to meet the demand for practical application.

In recent ten years, many algorithm of feature descriptors based on keypoints is widely applied, which adopt not only the gray level information of keypoints, also take the neighborhood information of keypoints to coping with scale zoom, rotation changes of image. From the data form characteristics of descriptor's component, local feature descriptor can be divided into two categories. One is based on the original feature's "absolute" magnitude; the second category is based on "comparison" binary bit strings.

2.1 LOCAL FEATURE DESCRIPTOR BASED ON "ABSOLUTE" MAGNITUDE OF THE ORIGINAL FEATURE

The most representative local feature descriptor based on "absolute" magnitude is SIFT (Scale Invariant Feature Transform) which proposed by David Lowe in 1999, and improved in 2004 [3, 4]. In the comparative experiment done by Mikolajczyk [5] on ten kinds of local descriptors including SIFT, and its extension algorithm has been proved to be robust with the strongest in the same descriptor type.

Overall, the SIFT descriptor has the following characteristics: (1) The SIFT feature is the local features of an image, not only has a good invariance to translation, rotation, scaling, illumination changes, occlusion and noise, but also maintain the stability of a certain degree to visual changes, affine transformation; (2) It is unique, informative, suitable for fast and accurate matching in mass feature database; (3) Even if few target also can produce a large number of SIFT feature vector; (4) scalability, it can be very convenient joint with other forms feature vector.

In the SIFT algorithm, to find out the multiscale extreme points in the image by constructing Gauss pyramid, as shown in Fig.1, this process consumes a large amount of computing and storage resources.

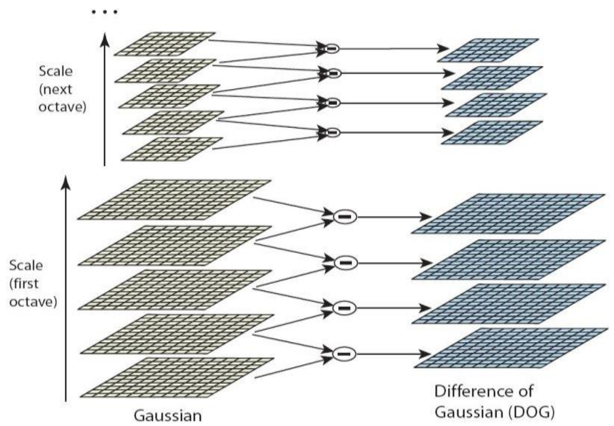

Fig.1. SIFT image pyramid constructing

SURF[6] (Speeded Up Robust Features) is an improved algorithm of SIFT, it uses integral image algorithm to replace the construction of Gauss pyramid in SIFT, and uses Hessian matrix which has advantages in speed and accuracy in detecting processing, thus the calculation efficiency is improved.

Wu Fuzhao etc., propose an image feature descriptor called Harris Related Features [7], the advantage of which cannot only construct feature point descriptor, but also straight line, curve descriptor. Guo Lisha, Li Junshan etc., proposed apply the FAST-9 operator to each layer in Gauss scale space images for extracting feature points, and then calculate the main direction in a circular area feature point as the center [8].

SIFT and SURF has the high quality of discriminate and intuition, but also with the high computational complexity due to the "absolute" magnitude of the original feature. So they still cannot get the good application on the high real-time situation.

2.2 LOCAL FEATURE DESCRIPTOR BASED ON "COMPARISON" BINARY BIT STRINGS

The representative local feature descriptor based on "comparison" binary bit strings is Brief (Binary Robust Independent Elementary Features), which is proposed by EPFL Calonder in ECCV2010 [9]. The algorithm selects some pairs of points near the feature point randomly, and then the gray values of a pair are compared and products the "0" or "1" combined into a binary bit strings as descriptor of the feature point. Each bit of a BRIEF descriptor are all obtained through two pixels randomly selected to do a binary comparison, considering the influence of noise, it is necessary to choose an appropriate Gauss kernel for image smoothing.

BRIEF has the advantage of speed, shortcomings in the: (1) have no rotation invariance; (2) sensitive to noise; (3) have no scale invariance. ORB (ORiented Brief Rublee) is proposed in ICCV2011 [10], solves the defects of above item (1), (2). In the ORB scheme, feature point detection operator directly use the FAST [11] which is a corner detection algorithm and famous for fast speed.

The FAST algorithm is used to detect whether the point P is a corner (in Fig.2 FAST9-16 for example). To the 16-point circle with P as the center and radius of 3, if at least consecutive 9 points in the circle are either sufficiently brighter or darker than the central point P, it can be confirmed that the P point is a corner and P is natural with rotation invariance.

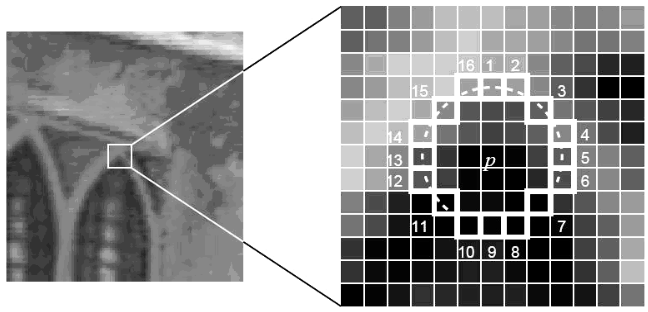

Fig.2. Diagram of FAST9-16 algorithm

In SIFT algorithm, the main direction of the feature points is determined by maximum and second maximum histogram bins' direction of the gradient histogram, the calculation is relatively large amount. In the ORB scheme, the main direction of the feature point is calculated by moment whose calculation is a small amount, and then based on the main direction to extract BRIEF descriptor.

The noise sensitive of BRIEF due to each bit construction of the descriptor string in pixel level. In ORB, does not use the comparison of pixel pairs, replaced by image blocks, namely contrast blocks' value summated by pixels in the block, pixel image block and value can be quickly computed by integral image.

According to statistics, the speed of ORB algorithm is the 100 times to SIFT, 10 times to surf [12]. This makes the ORB algorithm can meet for high real-time requirements. Despite the speed advantage, due to the characteristics of ORB with the FAST algorithm which has not the property of scale invariance, so ORB is still not resolved the scale invariance problem inherited from Brief.

BRISK (Binary Robust Invariant Scalable Keypoints) [13] solves the problem of scale invariance existed in ORB, meanwhile inherit its advantages. It introduces a scale space pyramid framework similar to the SIFT, so obtained feature points are more robust than ORB. The binary bit string of BRISK descriptor is also born of the difference of point pairs on neighbourhood of feature points. From the algorithm video demo [14] to see, the BRISK descriptor and has good performance on scale, rotation invariance and has fast speed for calculation, so it get the better application [15]. Similar to the SIFT algorithm, BRISK also need to construct image pyramid to get extreme points on multiscale, as shown in Fig.3.

In the BRISK framework, the scale-space pyramid layers consist of n octaves c_i and n intra-octaves d_i, for $i = \{0, 1,..., n-1\}$ and typically $n = 4$.

The octaves are formed by progressively half-sampling the original image (corresponding to c_0). Each intra-octave d_i is located in-between layers c_i and c_{i+1} (as illustrated in Fig.3). The first intra-octave d_0 is obtained by down sampling the original image c_0 by a factor of 1.5, while the rest of the intra-octave layers are derived by successive half sampling. Therefore, if t denotes scale then $t(c_i) = 2_i$ and $t(d_i) = 2_i \cdot 1.5$.

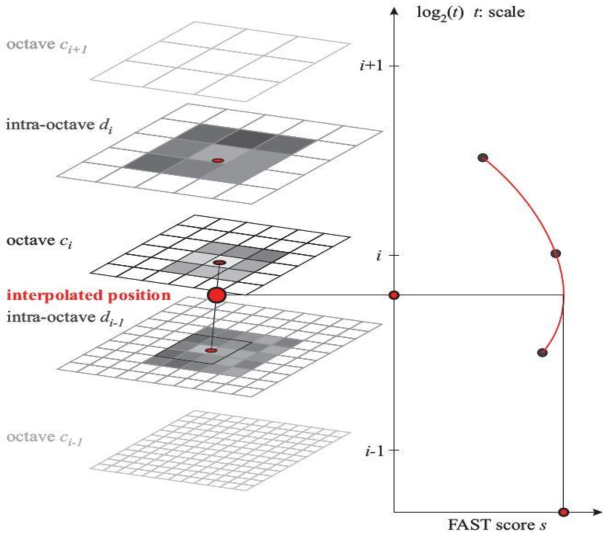

Fig.3. Scale-space extreme point's detection of Brisk

BRISK use the FAST 9-16 detector to extact feature points called keypoints. Initially, the FAST 9-16 detector is applied on each octave and intra-octave separately using the same threshold T to identify potential regions of interest. Next, the points belonging to these regions are subjected to a non-maxima suppression in scale-space: firstly, the point in question needs to fulfill the maximum condition with respect to its 8 neighboring FAST scores s in the same layer. The score s is defined as the maximum threshold still considering an image point a corner. Secondly, the scores in the layer above and below will need to be lower as well.

Considering image saliency as a continuous quantity not only across the image but also along the scale dimension, Brisk perform a sub-pixel and continuous scale refinement for each detected maximum. At first fitting a 2D quadratic function in the least-squares sense to each of the three scores-patches (as obtained in the layer of the keypoint, the one above, and the one below) resulting in three sub-pixel refined saliency maxima. Next, these refined scores are used to fit a 1D parabola along the scale axis yielding the final score estimate and scale estimate at its maximum. As a final step, re-interpolating the image coordinates between the patches in the layers next to the determined scale.

Scale-space pyramid layers and fit algorithm mentioned above will still consume a large number of computing resources and storage resources.

After given a set of keypoints, the BRISK descriptor is composed as a bit string of length 512 by concatenating the results of simple brightness comparison tests. The quantity of neighboring sample points of each keypoint is 60 in which produce 512 comparison pairs. The overall direction of keypoint is normalized as the sum of gray level gradient of 60 sample points. 60 sampling point template as shown in Fig.4.

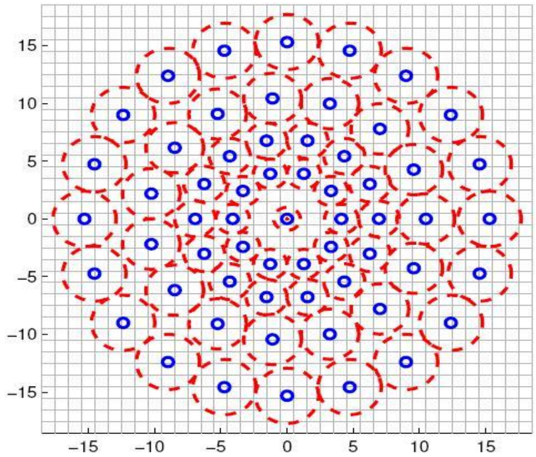

Fig.4. The BRISK sampling pattern with $N = 60$ points: the small blue circles denote the sampling locations; the bigger, red dashed circles are drawn at a radius σ corresponding to the standard deviation of the Gaussian kernel used to smooth the intensity values at the sampling points. The pattern shown applies to a scale of $t = 1$

Freak (Fast Retina Keypoint) [16] descriptor also has the characteristic of multiscale and rotation invariance. It based on the fact of human retinal cells distribution: center dense, surround sparse. The more close to the center, the area sampled more densely, and the surrounding area sampled sparse (keypoint template as shown in Fig.5), so it fits better the distribution characteristics of retinal ganglion cells.

In Fig.5, the retinal fovea called the macular region is the area of maximum visual resolution. Para is in the partial center area while Peri is in the surrounding area. Each red circle indicates a receptive field that's internal need to do Gauss smoothing.

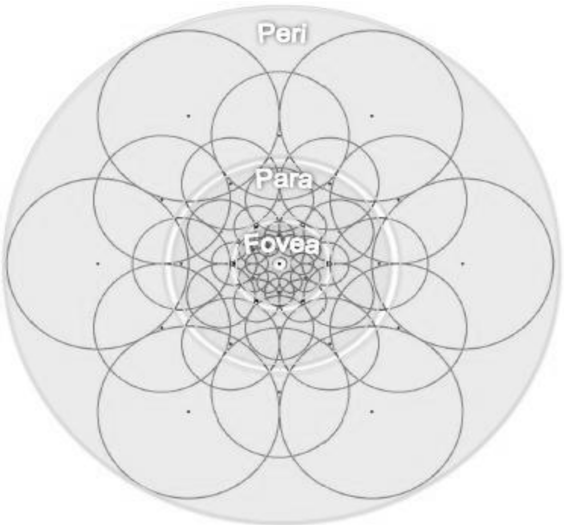

Fig.5. Illustration of the FREAK sampling pattern similar to the retinal ganglion cells distribution with their corresponding receptive fields. Each circle represents a receptive field where the image is smoothed with its corresponding Gaussian kernel

Feng Tang proposed OSID (ordinal spatial intensity distribution) algorithm [17] in CVPR2009, it is also a kind of local feature descriptor based on "comparison" and has stronger adaptability to illumination change compared to "absolute" based descriptors such as SIFT, SURF etc.

Tang Yonghe and Lu Huanzhang proposed the SCCH feature descriptor in the [18]. It extract Harris corner points of scale and rotation invariant as feature points in Gauss pyramid image, then the intensity difference between feature point and its neighborhood points is used as the feature descriptor.

All the "absolute" or "comparison" feature descriptors are artificial design directly. There are some researchers who use machine learning methods to get feature descriptor by data driven. This kind of feature descriptors include PCA-SIFT [19], Linear Discriminative Embedding [20], LDA-Hash [21] etc.

Whether it is the "absolute" or "comparison" based feature descriptor, they mostly use the feature points itself and the near field intensity and direction information. In fact, there is much information such as texture, color is not effective use. In [22], Zhang Xuewu, Ding Yanqiong et al. Make full use of the vision bionic mechanism, using the Gauss pyramid decomposition and Gabor pyramid decomposition of the image, extract the brightness, color and orientation features, achieves a high rate of accuracy in the application of copper strip surface defect detection.

In this research works mentioned all above, we can see that feature points of descriptor are usually multiscale extreme points. This is to guarantee feature is robust in scale change while multiscale extreme point extraction methods used currently are based on without exception pyramid decomposition method of various filter.

Based on the previous studies on "comparison" descriptors, the new extreme point extraction method presents in this paper.

3. DUAL MULTISCALE FAST ALGORITHM FOR FEATURE POINTS EXTRACTION

Classic local feature descriptors always extract extreme point of scale space by constructing the image pyramid, as shown in Fig.1 Brisk algorithm and Fig.3 SIFT algorithm. In original image pyramid decomposition, each sub image in one layer of pyramid is obtained from the last layer image by Gauss convolution and down sampling. The scale of lower image is the 1/2 to the upper image. This will produce a large number of sub images, need to consume large amounts of computing resources and storage resources, is not conducive to the development of actual product.

This paper proposes a new method, directly using corners extracted by FAST operator as the feature points of the image. But because the FAST operator itself only have the ability to detect corner point on current scale, and does not have the ability to detect multiscale corners, so the method "dual" to classical image pyramid is designed to achieve the multiscale FAST corner detection. The method is realized by multiscale modification of FAST operator, so have scale invariance ability for corner detection. In addition, the corners detected by common FAST operator have naturally rotation invariance, therefore scale extreme point extracted by proposed method also has rotation invariance.

The method proposed in this paper is described as follows:

1) The FAST9-16 operator as a reference, T(9-16) represent the detection threshold, then alter the FAST operator radius, can get FAST12-20, FAST14-24 increase with radius and FAST7-12, FAST5-8 operator decrease with radius, moreover calculate accordingly the detection thresholds in order T(12-20), T(14-24) and T(7-12), T(5-8). The detection threshold calculation method under various scales uses the Gauss convolution [23].
2) These FAST operators (total 5 from FAST5-8 to FAST14-24) are act on the original image, if there exist a point P which is a corner calculated by 4 of them, then the P is a keypoint with scale invariance, i.e. scale extreme point.
3) Gives the keypoint a main direction which can be calculated by moment method, the formula is as follows:

$$M_{ij} = \sum_{x}\sum_{y} x^i y^j I(x,y) \tag{1}$$

$$c_x = \frac{M_{10}}{M_{00}}, c_y = \frac{M_{01}}{M_{00}}, c_{ori} = \tan^{-1}\left(\frac{c_y}{c_x}\right) \tag{2}$$

The Eq.(1) calculates the moment, and c_{ori} in Eq.(2) is the main direction of keypoint.

The method proposed in this paper realizes extracting scale invariance corners by improving the FAST corner detection operator from multiscale ability. The method is "dual" to classic multiscale pyramid decomposition methods. It do not need decompose the original image by multiscale pyramid processing, only need to alter the operator scale. So this new method is named multiscale keypoints extraction based on Dual FAST algorithm.

This paper carried out a comparison experiment for verification of the proposed method. In Fig.6(a), the small circles denote the 176 keypoints extracted by proposed method, while in Fig.6(b) denote the 112 keypoints extracted by Brisk. There are 97 keypoints are coincident or only within 2 pixels deviation in two images. Keypoint of image(b) is less than keypoint of image(a) because keypoints extracted by Brisk need to meet more strict conditions.

From the comparison on efficiency of two algorithms, in the same computer configuration, (a) need 0.112 seconds and (b) need 1.357 seconds. The efficiency of proposed method shows 10 times efficiency to Brisk.

(a). Keypoints extracted by proposed method

(b). Keypoints extracted by Brisk

Fig.6. Keypoints extracted by proposed method comparison with Brisk

Thus, according to the comparison results, the keypoint extraction capability of proposed method are consistent with the Brisk algorithm, but hold a safe lead in efficiency, which proves the feasibility of proposed solution.

But the experiment is based on Brisk algorithm as the benchmark, although Brisk has good performance, but there have not been the strict mathematical derivation process. The method proposed in this paper also have the same problems. So the rationality of the proposed method shall be proved from the mathematical principles next step. It will have many problems to be solved.

4. CONCLUSION

This paper proposes a new method for multiscale extreme points extraction based on dual multiscale FAST algorithm which is different from the idea of classic pyramid decomposition. Experiments show that the intersection of keypoint sets extracted by proposed method and Brisk which uses pyramid decomposition is large, while the calculation of consumption of only Brisk 1/10, so it has better practical value.

ACKNOWLEDGMENT

This work is supported by the Open Fund for a Key-key “Information and Communication Engineering” Discipline of Zhejiang Province and the Ningbo Science Foundation Grant #2013A610113 to Hongwei Ying. We would like to thank Yongping Zhang, Gangyi Jiang for their helpful suggestions.

REFERENCES

[1] A. Mian, M. Bennamoun and R. Owens, “On the Repeatability and Quality of Keypoints for Local Feature-Based 3D Object Retrieval from Cluttered Scenes”, *International Journal of Computer Vision*, Vol. 89, No. 2-3, pp. 348-361, 2010.

[2] Guo Yu Lan, Lu Min, Tan Zhi-Guo and Wan Jian-Wei, “Survey of Local Feature Extraction on Range Images”, *Pattern Recognition and Artificial Intelligence*, Vol. 25, No. 5, pp. 783-791, 2012.

[3] David G. Lowe, “Object Recognition from Local Scale-invariant Features”, *Proceedings of the Seventh IEEE International Conference on Computer Vision*, Vol. 2, pp. 1150-1157, 1999.

[4] David G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91-110, 2004.

[5] Krystian Mikolajczyk and Cordelia Schmid, “A Performance Evaluation of Local Descriptors”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 10, pp. 1615-1630, 2005.

[6] Herbert Bay, Andreas Ess, Tinne Tuytelaars and Luc Van Gool, “Speeded-Up Robust Features (SURF)”, *Computer Vision and Image Understanding*, Vol. 110, No. 3, pp. 346-359, 2008.

[7] Wang Xuguang, Wang Zhiheng and Wu Fuchao, “Harris Correlation and Feature Matching”, *Pattern Recognition and Artificial Intelligence*, Vol. 22, No. 4, pp. 505-513, 2009.

[8] Guo Li-Sha, Li Jun-Shan, Zhu Ying-Hong and Ma Ying, “Fast Image Matching Algorithm Based on Multi-scale FAST-9”, *Computer Engineering*, Vol. 38, No. 12, pp. 208-210, 217, 2012.

[9] Michael Calonder, Vincent Lepetit, Christoph Strecha and Pascal Fua, “BRIEF: Binary Robust Independent Elementary Features”, *Proceedings of the 11th European Conference on Computer Vision: Part IV*, pp. 778-792, 2010.

[10] Ethan Rublee, Vincent Rabaud, Kurt Konolige and Gary

Bradski, "ORB: An Efficient Alternative to SIFT or SURF", *Proceedings of IEEE International Conference on Computer Vision*, pp. 2564-2571, 2011.

[11] Edward Rosten and Tom Drummond, "Machine Learning for High-Speed Corner Detection", *Proceedings of the 9th European conference on Computer Vision –Part I*, pp. 430-443, 2006.

[12] Ren Jie, Zhou Yu, Yu Yao, Du Si-dan and Wang Zi-qiang, "Real-time augmented reality based on ORB", *Application Research of Computers*, Vol. 29, No. 9, pp. 3594-3596, 2012.

[13] Stefan Leutenegger, Margarita Chli and Roland Y. Siegwart, "BRISK: Binary Robust Invariant Scalable Keypoints", *Proceedings of IEEE International Conference on Computer Vision*, pp. 2548-2555, 2011.

[14] ETH, Autonomous Systems Lab, Demo Video [CP/OL], http://www.asl.ethz.ch/people/lestefan/personal/BRISK.

[15] Xu Yun-xi, Jiang Yun-liang and Chen Fang, "Rapid scene matching algorithm based on Brisk for inertial integrated navigation system", *Journal of Optoelectronics Laser*, Vol. 23, No. 8, pp. 1589-1596, 2012.

[16] Alexandre Alahi, Raphael Ortiz and Pierre Vandergheynst, "FREAK: Fast Retina Keypoint", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 510-517, 2012.

[17] Feng Tang, Suk Hwan Lim, Nelson L. Chang and Hai Tao, "A Novel Feature Descriptor Invariant to Complex Brightness Changes", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2631-2638, 2009.

[18] Tang Yong-he, Lu Huan-zhang and Hu Mou-fa, "An Image Matching Algorithm Based on SCCH Feature Descriptor", *Journal of Electronics and Information Technology*, Vol. 33, No. 9, pp. 2152-2157, 2011.

[19] Yan Ke and Rahul Sukthankar, "A More Distinctive Representation for Local Image Descriptors", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. II-506 -II-513, 2004.

[20] Bo Li, Chun-Hou Zheng and De-Shuang Huang, "Locally Linear Discriminant Embedding: An Efficient Method for Face Recognition", *Pattern Recognition*, Vol. 41, No. 12, pp. 3813-3821, 2008.

[21] Christoph Strecha, Alexander M. Bronstein, Michael M. Bronstein and Pascal Fua, "LDAHash: Improved Matching with Smaller Descriptors", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 34, No. 1, pp. 66-78, 2012.

[22] Zhang Xuewu, Ding Yanqiong, Duan Dunqin, Gong Fang, Xu Lizhong and Shi Aiye, "Surface defects inspection of copper strips based on vision bionics", *Journal of Image and Graphics*, Vol. 16, No. 4, pp. 593-599, 2011.

[23] Zhang Ke, Wang Qi and Wang Li, "New construction algorithm for convolution mask in scale space inspired by mechanism of retina", *Opto-Electronic Engineering*, Vol. 33, No. 2, pp. 76-80, 2006.