

Video Object Matching Based on SIFT and Rotation Invariant LBP

Deng Yi^{1,2}, Lu Jianguo³, Qu Xilong^{*2}

¹School of Mechanical and Electrical Engineering, Central South University, Changsha, China

²College of Mechanical Engineering, Hunan Institute of Engineering, Xiangtan, China

³Pre-school Education College, Jiaozuo Teachers College, Jiaozuo, China

*Corresponding author, e-mail: zshaob@163.com

Abstract

Object detection and tracking is an essential preliminary task in event analysis systems (e.g. Visual surveillance). Typically objects are extracted and tagged, forming representative tracks of their activity. Tagging is usually performed by probabilistic data association. However, as data may have been collected at different times or in different locations, it is often impossible to establish such associations in systems capturing disjoint areas. In this case, appearance matching is a valuable aid. This paper proposes a object matching method for multi-camera by combining HOG and block LBP, and computes accuracy rate by SVM. Using independent tracks of 30 different persons, we show that the proposed representation effectively discriminates visual object and that it presents high resilience to incorrect object segmentation and illumination. Experimental results show that the average accuracy rate is up to 80%.

Keywords: object matching, multi-camera, non-overlapping views, HOG, LBP

Copyright © 2013 Universitas Ahmad Dahlan. All rights reserved.

1. Introduction

The object tracking under the environment of non-overlapping multi-camera surveillance is one of the crucial issues in the wide area video surveillance studies. Compared with single-camera and overlapping multi-camera tracking, non-overlapping multi-camera tracking has its own particularity. First of all, in non-overlapping camera views, object image are not continuous in spaces and time; Secondly, the appearance model of a object varies under in different views because of the influence of illumination, posture and the camera's parameters. Therefore, most of the traditional video object tracking techniques and researching frameworks are necessary to be improved.

The current solution is extract the external feature and the spatiotemporal feature from the object, and combined with the constraints of time and spaces when transferred between cameras with the help of topology relations, and then matching then feature by using correlation algorithm on different camera so as to gain the corresponding relationship of the object in different cameras, achieve the object. The extracting and matching of the correlation feature of the object is the key point of the study, matching effect will directly affect the performance of the tracking algorithm. At present, the matching method mainly includes the establishment of appearance of the object (Appearance Model), transfer model between the appearance of the object in different cameras (Transfer Model), and the estimation of similarity degree, etc. Appearance model is established by using the object feature captured in the different cameras which includes color, point, line, regional, geometric, feature model etc. In the existing research results, researchers often choose one or combine some of these features to build the appearance model matching.

Researchers have tried using a variety of appearance feature of human object matching which mainly based on the color feature [1-6], point feature [7] or the combination of them at home and abroad[8]. The color-based mainly uses the color histogram, and combined with the brightness variation function to respond to changes in brightness between different cameras. The advantage adapt for the non-rigid object matching ,things always has two sides followed by disadvantages: one is light conditions, another is the BTF acquisition between cameras, the overhead computational cost leads it is difficult to meet the real-time requirements, so this paper focus on the LBP feature under blocked Gaussian [9-11]. LBP descriptors can maintain good

performance though big changes of illumination on the object's surface. The original LBP operator is relatively large influenced by the noise, however, it is improved by the blocked Gaussian differential scale description, that is, robust to noise, maintaining stable illumination at the same time, to the color feature, can be easily response to illumination changes, without calculating the BTF. While the point-based, such as the scale-invariant feature, visual words need to be computed, that is cluster unfixed points according to the clustering algorithm, to generate the object model, the sixth reference reached 11110 cluster centers, so the advantage of HOG feature is self-evident [12-15]. HOG feature and SIFT feature are similar [16-17], but no descriptor number difference, so HOG feature are robust against registration error in that it compute the gradient direction which located in small area of the image, and the HOG computed feature facilitate extraction of the fixed dimension, does not involve the point number control [18-19]. The parameter extracted from adjusting process of HOG feature can effectively describe the shape information of the human body, to quantify the position and direction of spaces, to some extent, restrain the translational and rotational effects [20]. The illumination normalization process in the HOG computation, and the LBP descriptor invariant feature of illumination, combined such this combination makes the feature have a strong robustness to illumination changes.

2. Object Detection and Tracking under the Single-Camera

The first operation is segmentation of the moving object using background subtraction techniques, and then is object tracking using particle filter for the video sequences from each camera.

The background differential method, inter-frame differential method and optical flow method are available for moving object segmentation. This paper presents the background differential method based on the statistical background subtraction techniques. The advantage of this method is to reconstruct the background successfully whether the background of image is complete or not. First, put the N frames of video in the buffer area, transformed into grayscale images and count the pixels number of the same position, select the highest as the position's pixel values so as to construct grayscale background. Then reconstruct the colorful background according to the corresponding relation between gray pixel and the original colorful image. Next, segment the object binary image by using the current frame subtracts the background frame. Finally is tracking using the particle filter method after segmentation of moving object. In the test, the particle number N is set to 500, the width and height of object external rectangle are respectively 20 and 44. The tracking results are shown in figure 1.

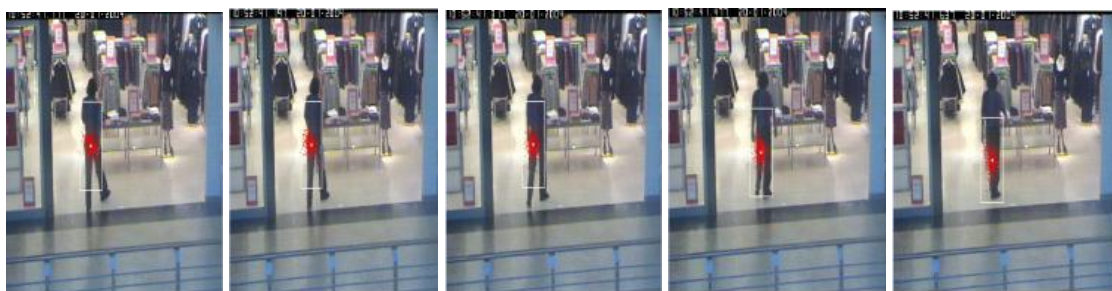


Figure 1. Particle filter tracking results

3. Matching Algorithm Based on the HOG and Blocked Gaussian Differential LBP

3.1. HOG Feature

The HOG (Histograms of Oriented Gradients) feature is a local descriptor which commonly used as the descriptor for object detection. Following is the five basic of feature extraction:

- 1) Ignoring the image normalization process, to transform input image into grayscale image;

- 2) Computing the gradient from the image's horizontal ordinate and vertical ordinate, then further, computing the matrix of gradient direction of each pixel position.
- 3) Dividing the image window into several small areas of same size which called "cells". Then operate on all pixels in each cell, count histogram of gradient direction from the intracellular pixel in accordance with the previous fixed direction division standard.
- 4) Normalizing gradient histogram in the block. Different blocks sharing a same cell and cell's normalization based on the different blocks so as to different computing results. Therefore, the feature of same cell will appear different results many times in final result's vector.
- 5) Collecting HOG feature. Collect computed HOG feature from all the overlapping blocks and combine them into the final feature vector in the detection window.

This paper omitting the normalization process of the object image and segment the object's size without the mandatory conversion. And divide object into cells at the size of 5*5 pixel per one, four (2*2) cells composed of a block, so multiple overlapping blocks covering the entire object area. Since the computing of histogram in accordance with the 180 degree of gradient direction, and divided into 9 parts, the dimension of is nine in each cell's gradient histogram, and each block consists of four cells, so 4*9=36 dimensional feature vectors compose a block. In the test, the scanning sequence of the block is line-by-line and the interval is two, namely the odd pixel one as the starting pixel of block. Serializing feature vectors of the first 36 blocks when collecting the HOG feature, and composing the final feature vector of object images. Because the dimension of the feature of each block is 36, therefore, a total of 36*36=1296 dimensions feature vectors for each image.

3.2. Blocked Gaussian Differential Scale LBP feature

LBP is a texture descriptor which portraying local gray change of the image using local texture mode and count the local texture structure according to the texture spectrum histogram. The basic LBP feature is a 3 × 3 neighborhood binary description. 3 × 3 neighborhood taken from any point of the original image, and the point's pixel value is the threshold which compared respectively with the gray value of the eight pixels in neighborhood, and generate a binary mode which encoded by the binary pattern. The value of the LBP feature of the pixel of center point is equal to value that convert binary code into a decimal number.

This paper describes the LBP feature of blocked Gaussian scale spaces of image, the basic steps: first setting up different, Gaussian scale spaces for the object image after segmentation, getting the image of the Gaussian differential scale spaces according to the difference operation which be used for the computing of histogram of the blocked LBP feature, and serializing the as the histogram of the object image. In short, Two basic steps: 1) Setting up the Gaussian differential scale spaces 2) Computing histograms of the blocked LBP.

1) Constructing Gaussian differential scale spaces

Generally speaking, the purpose of constructing this scale spaces is to simulate the multi-scale feature of image data. In the test, Gaussian function is used to establish the scale spaces, because Gaussian convolution kernel is the only linear kernel that can help to realize scale transformation. Here formula [1] & [2] is used to define the scale spaces of double-dimensional image and Gaussian function:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad [1]$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad [2]$$

To lessen the complexity of time and spaces, the concept of the Differential scale spaces of Gauss (DOG), generated from different scales of Gaussian differential kernel and image convolution is introduced here. The test set two different scale coordinates to create two different scale spaces, where the scale factors were respectively 2 and 0.08. Then the Gaussian differential scale spaces are gained by the differential computing. Following the DOG computational formula:

$$D(x, y, \sigma) = (G(x, y, \sigma_1) - G(x, y, \sigma_2)) * I(x, y) = L(x, y, \sigma_1) - L(x, y, \sigma_2) \quad [3]$$

2) Blocked LBP histogram

Blocked LBP feature describe the object by multi-section histogram sequence of images with LBP feature under differential scale spaces. First, divide the Gaussian differential images in to sub-blocks by the size of $2 * 2$ pixels (Ensuring that all sub-block do not overlap each other), and determine the values of all points in each sub-block. Second, analyze each LBP sub-block with histograms, then joint all the histogram gained a final sequence as the LBP description. To extract this LBP descriptor, it set factors as: $P=8$, $R=2$, i.e., the description is fulfilled by 8 points circle the center pixel with 2 as the radius. So the dimension of LBP descriptor is $256 * 4=1024$.

3.3. Matching the Implementation Parameters of the Moving Object

This study selectively uses the support vector machine (SVM) which shows unique advantages in solving small sample, non-linear and high dimensional pattern recognition. The basic idea is, using a nonlinear mapping p , to map the sample spaces to a high dimensional, and even infinite-dimensional feature spaces (Hilbert spaces), so that making the nonlinear separable problem in original sample spaces into a linear separable one in the feature spaces.

The test were conducted on the platform of Matlab R2010a, whose configuration is Pentium (R) Dual-Core CPU 2.20GHz RAM 2.00GB. SVM classification follows the standards by Mr. Lin Zhiren in Version libsvm-3.1. The HOG feature extracted from the samples and LBP feature under blocked Gaussian differential scales were as the input of SVM. Before the classification, in order to balance the function of each feature component of a feature vectors works, external normalization processing was to take on the feature vectors by the following normalizing mapping:

$$f : x \rightarrow y = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad [4]$$

Wherein x , y , respectively, stands for sample feature and post-normalization sample feature, $x_{\min} = \min(x)$, $x_{\max} = \max(x)$. The effect of normalization is that the original data is structured to $[0,1]$ range, i.e. $y \in [0,1]$. This process is called Range $[0,1]$ normalization. Test take radial basis function as kernel function, then gaining penalty parameter c and radial basis-kernel function parameters g through cross-validation.

4. Test Results and Analysis

4.1. Data Sets

In order to evaluate the performance of the proposed method, we created a data set: D_{30} . D_{30} includes 30 independent visual objects from Shopping Center data set of CAVIAR Test Case Scenarios. These 30 moving object is segmented and selected from 26 video sequences in the original data set by the method of object detection introduced in Section II. 3300 frame sequence is used for training and testing in the experiment, and this frame sequence includes the 30 independent objects, as is shown in Figure 2. The Experiment is conducted in 3 separate groups, each of whom includes 10 objects. Test are repeated in different time periods, at different places on a same object with the following two conditions taken into consideration: First, due to the environment the two cameras locate is different, results in light changes in different sequences; second, due to the angles of the two cameras are different, with the distance change of the moving object in the sequence, leads to some parts of the object being out of monition, as is shown in Figure 3 and Figure 4.

4.2. SVM Classification Results

The paper extracted HOG feature and LBP feature under blocked Gaussian differential scale as the input of SVM, and experimented in 3 categorized object groups. Group1 selected 50 frames as the training samples and 90 as the testing samples, while group 2 and group 3 selected 50frames as the training samples and 50 as the test samples. Illumination changes and partially blocked object in the video frames are included in the samples, which are used in the calculation of matching accuracy under the change in light and partially blocked cases.



Figure 2. The 30 available visual objects in dataset

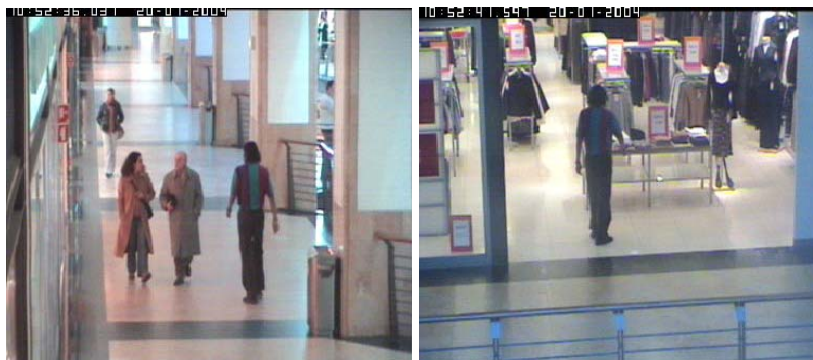


Figure 3. The illumination of same object changes in the video frames in the different cameras and time

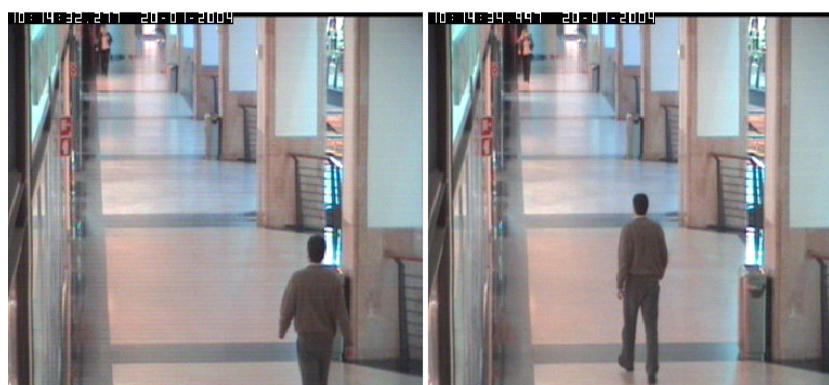


Figure 4. The appearance of same object changes in the video frames in the one camera at different time

The LBP block size is 2×2 , the Gaussian differential scale σ were respectively 2 and 0.08.

First, conduct the test separately all by the algorithm afore-proposed in 3 categorized groups. Figure 5(a) shows the categorized confusion matrix data of Group 1 based on the algorithm afore-proposed, in which, the $bestc=0.00097$, $bestg=0.0078$, and the matching accuracy is 91.1%. The test the three groups in test were repeated five times, the average

accuracy of the experimental results is listed in Table 1. As the table shows, the object image to object matching accuracy is all over 80%.

Table 1. The accuracy of the proposed algorithm matching

	Group 1	Group2	Group 3
Accuracy rate	85.6%	82.1%	80%

Next, categorize the three groups by HOG feature and LBP feature. Figure 5 (b) shows the confusion matrix data of Group 1 categorized only by HOG feature, in which, the bestc=2, bestg=0.03 and the matching accuracy was 72.2%. The experimental result of the 3 groups is the average value of 5 repetitions, as is shown in Table 2. Figure 5 (c) shows the confusion matrix data of Group 1 categorized only by LBP feature, in which, the bestc=0.00097, bestg=0.003 and the matching accuracy was 76.6%. Shown in Table 3 is the average value of the three groups' matching accuracy after 5 repetitions.

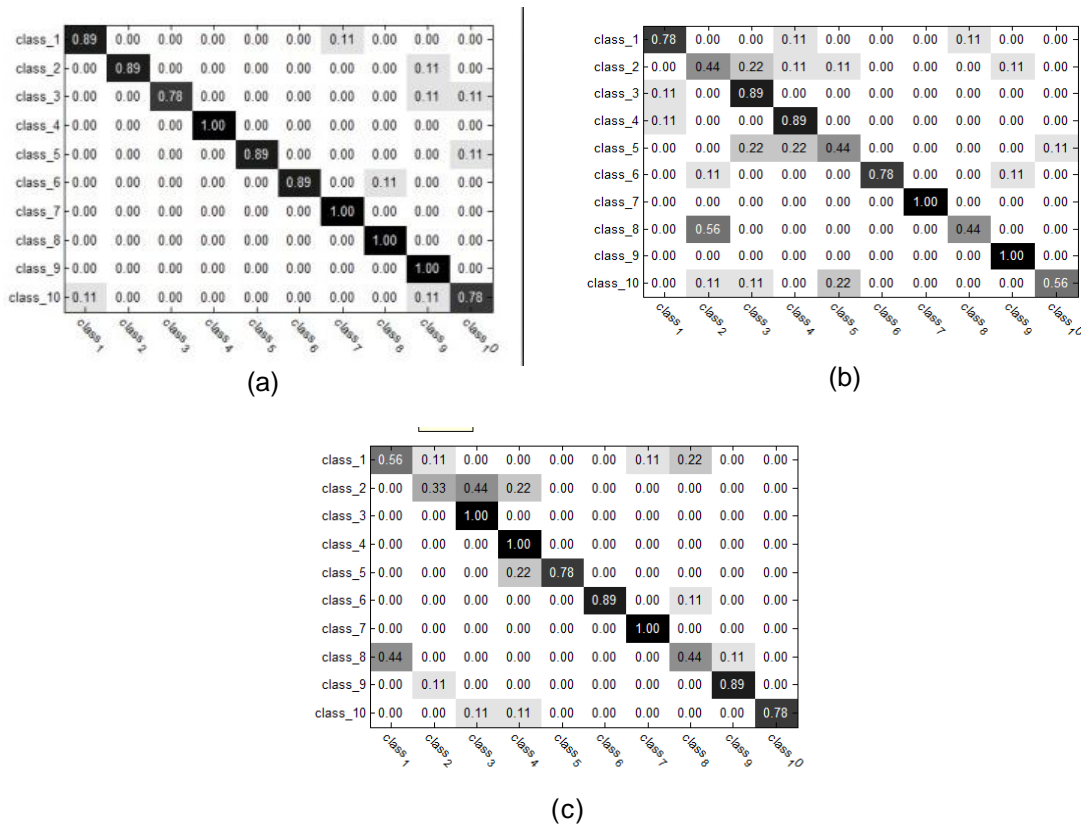


Figure 5. Confusion matrix

Table 2. The accuracy rate under single-HOG matching algorithm

	Group 1	Group 2	Group 3
Accuracy rate	74.4%	70%	74%

Table 3. The accuracy rate under single-LBP matching algorithm

	Group 1	Group 2	Group 3
Accuracy rate	68.8%	70%	72%

From the above three tables, it can be concluded that, due to the robustness HOG feature showed to the objects' posture, it can capture the objects regardless of their slight movements, while since the blocked the LBP feature under Gaussian scale difference showed better robustness to noise, it can play a complementary role to HOG-feature matching in complex background changes of light. Seeing from the confusion matrix A, when matching by the proposed algorithm feature, it may lead to mismatching due to the objects' changes of postures and the similarity in outer skeleton. From the confusion matrix A, B and C, for each kind of the objects, the matching accuracy rate is higher with the two complementary feature combined than with them separated, to be exact, the former is about 10%-20% higher than the latter. Namely, the effect when monitoring with the two feature combined is better than that when doing so alone with any of them, HOG feature or LBP feature.

Different solutions are oriented to different assumptions, different geometric constraint and camera topological relation constraint, for example. To this extent, it's kind of unrealistic to compare the method propose in this paper with that proposed in works [6] and [7] directly. However, what can be compared directly is the matching performance of SIFT and color spectral methods adopted in works [6] & [7]. Table 4 shows classification accuracy comparison. And from it, We know that when matching with SIFT feature, the data of its descriptor is enormous, this can also be seen from the feature's size. As for accuracy, the feature proposed in this paper is the highest among the three methods. And see from feature's size and classification rate comprehensively, the proposed method achieved very good results.

Table 4. Comparison of accuracy rate

Feature	Classification rate	Feature size
MCSHR	71.6	Variable from ~200
SIFT	67.2	Variable from 1000 to 5000
HOG-LBP	85.6	Fixed 2536

5. Conclusion

The visualized multi-cameras surveillance system mainly depends on the process of object segmentation and tracking. So far, most tracking technology is based on single-camera tracking, and many algorithms have been proposed for multi-camera tracking, but usually, the correlation between cameras are known. The object matching algorithm of performance model which combined Histograms of Oriented Gradients (HOG) feature and blocked Gaussian differential scale LBP feature, brought forth in this paper, proven to be effective in the non-overlapping views multi-camera tracking system. The simulations reveal that: the multi-camera object matching algorithm combined HOG feature and LBP feature under blocked Gaussian differential scale show robust performance in the situation where illumination changes and small part of covering. The combination of the two complementary feature make the matching performance enhanced, and meet well the needs brought about by illumination changes and partial covering. Aided tracking in multi-camera tracking system performs, for it is qualified in object distinction and adapted to the changes of object.

Acknowledgements

The authors are grateful to the anonymous reviewers. Authors gratefully acknowledge the projects supported by Provincial Natural Science Foundation of Hunan (13JJ9022 & 11JJ9018) and the project supported by Provincial Science & Technology plan project of Hunan (2012GK3077) for supporting this research.

References

- [1] Javed O, Shafique K. Modeling inter-camera spaces-time and appearance relationships for tracking across non-overlapping views. *Computer Vision and Image Understanding*. 2008; 109(2): 146-162.
- [2] Cai Y, Chen W, Huang K, Tan T. Continuously tracking objects across multiple widely separated cameras. *LNCS*. 2007; 4843: 843-852.

- [3] Jeong K, Jaynes C. Object matching in disjoint cameras using a color transfer approach. *Machine Vision and Applications*. 2008; 19(5): 88-96.
- [4] Cheng ED, Piceardi M. Disjoint track matching based on a major color spectrum histogram representation. *Optical Engineering*. 2007; 46(4): 1-14.
- [5] Gilbert A, Bowden R. Incremental, scalable tracking of objects inter camera. *Computer Vision and Image Understanding*. 2008; 111: 43-58.
- [6] Teixeira LF, Real LC. Video object matching across multiple independent views using local descriptors and adaptive learning. *Pattern Recognition Letters*. 2009; 30(2):157-167.
- [7] Cheng ED, Piccardi M. *Matching of objects moving across disjoint cameras*. IEEE International Conference on Image Processing, Atlanta. 2006; 1769-1772.
- [8] Patwardhan KA, Sapiro G, Morellas V. *A graph-based foreground representation and ITS application in example based people matching in video*. IEEE Intl. Conference on Image Processing. 2007: 37-40.
- [9] Huang Qian, Liu Jun, Peng Xiao Sun Tao pedestrian detection based on local binary pattern feature. *Computer Engineering and Design*. 2011; 32(6): 2119-2123.
- [10] Sun N, Ji Z, Zou C. Gender classification based on local binary pattern. *Journal of Huazhong University of Science and Technology: Nature Science Edition*. 2007; 35: 177-181.
- [11] Zheng Yongbin, Huang Xinsheng, Feng Songjiang. SIFT and rotation invariant LBP combined image matching algorithm. *Aided Design and Computer Graphics*. 2010; 22 (2): 286-291.
- [12] Dalal N. Histograms of oriented gradients for human detection. *CVPR*. 2005; 886-893
- [13] Zhou Jinzhi, Juan. Based HOG human detection method improved software GUIDE. 2011; 10(4): 76-78.
- [14] LIU Qing, Guo Zhigang, Guo Jianming, Li Longli convert gradient direction histogram viewing angle and rotation angle changes. *Control Theory and Applications*, 2010; 27(9): 1269-1272
- [15] Sails Wen, Qu Zhen deep, Yan Ji. *Moving object based on CKPCA-HOG and support vector machine classification algorithm*. Harbin Engineering University. 2011; 32(5): 643-649.
- [16] Amin Mohamed Ahsan and Dzulkifli Bin Mohamad. Features Extraction for Object Detection Based on Interest Point *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013; 11(5).
- [17] Yongmei Zhang and Li Ma and Rui Zhang. A Quick Image Registration Algorithm Based on Delaunay Triangulation. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013; 11(2): 761-773.
- [18] Lipton A, Fujiyoshi H, Patil R. *Moving Object Classification and Tracking from Real-Time Video*. Proceedings of IEEE Workshop on Applications of Computer Vision, Princeton NJ. 1998; 8-14.
- [19] Antonio F, José CC, Javier MC, Rafael MT. Optical flow or image subtraction in human detection from infrared camera on mobile robot. *Robotics and Autonomous Systems*. 2010; 58(12): 1273-1281
- [20] NIU Chang-feng, Jointly, Liu, Yushu. Object tracking method based on SIFT feature and particle filter. *Robot*. 2010; 32 (2): 241-247.
- [21] Yan Y, Wang J, Li C, Wu Z. *Object tracking using SIFT feature in a particle filter*. IEEE 3rd International Conference on Communication Software and Network. 2011; 384-388.
- [22] Qu Xilong, Hao Zhongxiao, Bai Linfeng, Research of Distributed Software Resource Sharing in Cloud Manufacturing System. *IJACT: International Journal of Advancements in Computing Technology*. 2011; 3(10): 99-106.