Object Recognition and Tracking based on Object Feature Extracting

Yong-Hwan Lee[1], Hyochang Ahn[2*], Han-Jin Cho[1], and June-Hwan Lee[1]
[1]Far East University, Eumseong Chungbuk, 27601, Republic of Korea
hwany1458@empal.com, hanjincho@hotmail.com, rainbow@kdu.ac.kr
[2]Dankook University, Yongin-si, Gyeonggi-do, 448-701, Republic of Korea
youcu92@dankook.ac.kr

Abstract

Object recognition and tracking is one of the most important task in the field of computer vision and surveillance system. Among many proposed, object tracking using a feature matching method is popular and accurate, however, it has a room to improve on high computational complexity and weak robustness in various environments. This paper proposes a robust object recognition and tracking method, which uses an advanced feature matching for the use of real time environment. The proposed method enables to recognize an object using invariant features, with reducing the dimension of feature descriptor, compared to the existing algorithm. The experimental results show that the proposed recognition and tracking method outperforms the conventional tracking approaches in terms of tracking accuracy and computing time.

Keywords: Object Recognition, Object Tracking, Feature-Matching Approach, Feature Extraction

1 Introduction

In recent years, video surveillance and security monitoring systems have developed rapidly to monitor several circumstances of the public area, as these systems have great powered performance and precision. Several surveillance cameras have been installed for the attention and necessity of the security and surveillance, respectively. Most of object tracking approaches based on feature matching have a problem, showing high computational complexity and/or weak robustness in various environments.

To efficiently track a dynamic object in a video sequence, at first, feature points are extracted from the interest object. The extracted features then recognize the target object, and the detected and recognized object is continuously tracked on the input stream [22, 21, 6]. It is an important technology in computer vision. The object recognition consists of two main steps. One is extracting an interest feature point in a target object. Another is matching a corresponding point at a target video sequence. In object recognition based on feature, extraction of accurate feature in the target has influenced the performance of the object recognition. The performance of recognition can be improved by a large number of feature points, which are the extracted interest region. However, it is impossible to recognize the object in real time environment owing to increased computing complexity. On the other hand, it cannot carry out precise recognition. Thus, the algorithm of object recognition requires detecting accurate object and decreasing computational complexity in real time surveillance system. One of the common approaches using an interest object tracking is CamShift [19, 23, 24], which is an object tracking method by color histogram as its target model [14, 7]. In this paper, we propose robust object recognition and tracking scheme using advanced feature matching for real time environment.

Journal of Internet Services and Information Security (JISIS), volume: 5, number: 3 (August 2015), pp. 48-57

*Corresponding author: Department of Applied Computer Engineering, Dankook University, 152, Jukjeon-ro, Suji-gu, Yongin-si, Gyeonggi-do, 16890, Republic of Korea, Tel: +82-(0)10-3016-8792

The rest of this paper is organized as follows. Section 2 gives a short survey of some related works for feature extraction. Section 3 concentrates on the proposed method, which uses the CamShift for object tracking and utilizes a novel advanced feature extraction method for object recognition. Prototype system is implemented and the experimental result for evaluation is shown in Section 4. Finally, Section 5 presents conclusions and future work.

2 Related Works

Video surveillance system is mainly composed of object recognition and tracking [2, 16]. In object recognition, the most important thing is accurate feature extraction. A good feature descriptor must have the ability to handle intensity, rotation, scale, and affine variations. Feature extraction methods are generally used scale invariant feature transform(SIFT) and speeded-up robust features(SURF) in object recognition [11, 12, 9, 13]. SIFT aims to resolve the practical problems in low level feature extraction and their use in matching images. SIFT is a method of transformation into a set of regional characteristics with robust feature against the size, rotation, and projection of the object [11]. SIFT method as image size increases will be plenty to calculate the amount of data because of the high-dimension characteristic [15, 3]. If the amount of data is increased, the method has the problem which increased the amount of computation time as increasing volume of calculation [17]. Features are located using an approximation to the determinant of the Hessian matrix in SURF algorithm [6, 1]. It is used due to its stability and repeatability, as well as its speed. The Hessian is constructed by an ideal filter. It convolves the input image with the second-order derivatives of a Gaussian of a given scale. SURF uses the integral image concept. SURF has fast feature extraction and feature descriptors to reduce the complexity of the operation in feature extraction and matching step on SIFT method [19]. It performs good result and high speed using feature extraction methods and feature descriptors by decreasing processing time. However, SURF has plenty of computational time for object recognition in real time. Thus, we need to develop an algorithm to perform precise object recognition while reducing the amount of computational time. Many tracking algorithms have been proposed in earlier researches.

In a tracking video sequence, an object can be defined as anything that is interesting for analysis [5, 10, 20]. Object tracking methods generally use MeanShift and CamShift [25, 14]. The MeanShift is a kind of tracking algorithm based on external features, with which real-time tracking for non-rigid object can be realized [14]. The MeanShift algorithm is an efficient approach to tracking objects whose appearance is defined by histograms. In the MeanShift procedure, the points in the d-dimensional feature space as an empirical probability density function where dense regions in the feature space correspond to the local maxima or modes of the underlying distribution. For each data point in the feature space, one performs a gradient ascent procedure on the local estimated density until convergence [21, 3]. The CamShift is one of the most important algorithms for object tracking [4, 8]. The CamShift is an adaptation of the Mean Shift algorithm in computer vision. The primary difference between CamShift and MeanShift algorithm is that CamShift uses continuously adaptive probability distributions while MeanShift is based on static distributions, which are not updated unless the target experiences significant change in shape.

3 Proposed Method

CamShift is a common algorithm for tracking an interest object for real time environments. However, the algorithm uses only color features, so it is not robust to the surrounding environment and illumination. This method has the problem of losing the interest target when a similar color exists in the background because it is sensitive to illumination and noise. This problem can be solved by the integrated method which recognizes the target object using SURF for feature extraction and can track the object by CamShift. Since SURF has the advantage which can be found in invariant feature points of rotation and scaling, it can recognize the same object regardless of the angle and distance of the camera. SURF can find out the target object using feature extracting and matching. Fast and accurate object recognition needs to find out matching points efficiently in real time environment. Although SURF has low computational complexity less than SIFT method, it is not suitable for recognizing objects in real time.

Therefore, the proposed method extracts the features and finds out the corresponding matching points in video sequence. Also the computational complexity is reduced to efficiently decrease dimension in feature descriptor to carry out object recognition. The proposed object tracking system has a simple architecture which is based on a set of cyclically interconnected modules. Each module deals with a specific type of input data that is elaborated to provide appropriate data to the next module. Figure 1 describes diagram of the proposed object recognition and tracking system using feature extraction.

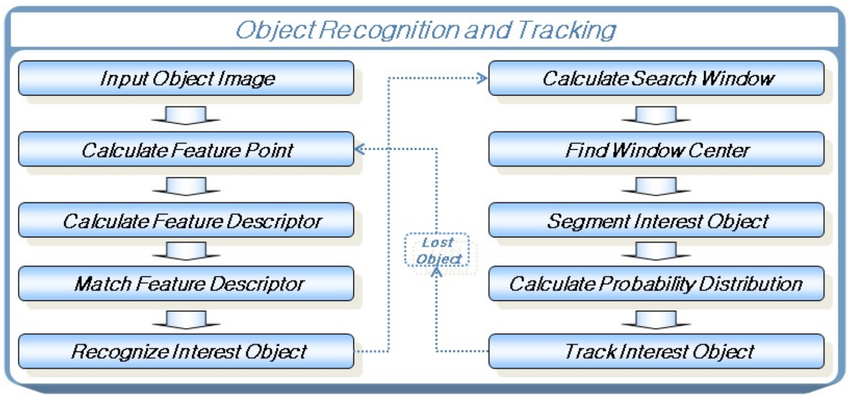


Figure 1: Diagram of the Proposed Object Recognition and Tracking System.

3.1 Object Recognition

Speeded Up Robust Feature (SURF) is a robust illumination changes and invariant scale detector and descriptor for feature point. The integral image is used to SURF for decrement of computation. The sum of all pixels in the selected partial region is calculated by only performing four operations. Therefore, when scale space is generated, the amount of computational time is reduced [3, 17].

The next step of object recognition is quickly extracting features using the extractor based on an approximation of Hessian matrix in interest points. In this case, the extractor extracts the features of images for changing of various scales by resizing the box filter without changing the image scale. Figure 2 shows image pyramid and box filter for extraction of the features.

Hessian matrix can be obtained by convolution of the second derivative of Gaussian filter and an image, and it can be expressed as Equations 1 and 2 [1].

$$H(X,\sigma) = \begin{bmatrix} L_{xx}(X,\sigma) & L_{xy}(X,\sigma) \\ L_{xy}(X,\sigma) & L_{yy}(X,\sigma) \end{bmatrix} \tag{1}$$

$$L_{xx}(X,\sigma) = I(x,y) * \frac{\partial^2}{\partial x^2} g(\sigma) \tag{2}$$

where $L_{xx}(X,\sigma)$ denotes the convolution of the second derivative of Gaussian filter and an input image at the point of $X = I(x,y)$ in an input image having a scale of σ. In addition, $L_{xy}(X,\sigma)$ and $L_{yy}(X,\sigma)$ are the represented convolution of the second derivative of Gaussian filter and an input image for xy

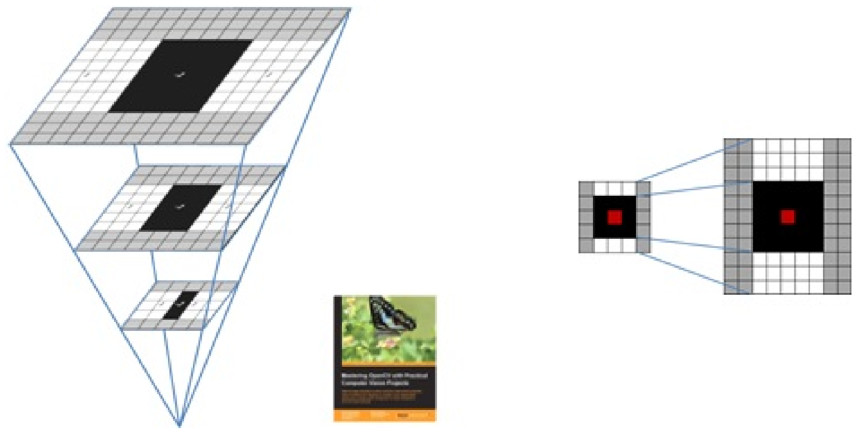


Figure 2: Image pyramid and Box filter.

direction (diagonal) and y direction (vertical). The method uses the box filter adapted approximation of convolution of the second derivative of Gaussian to solve the problem of increasing processing time [18].

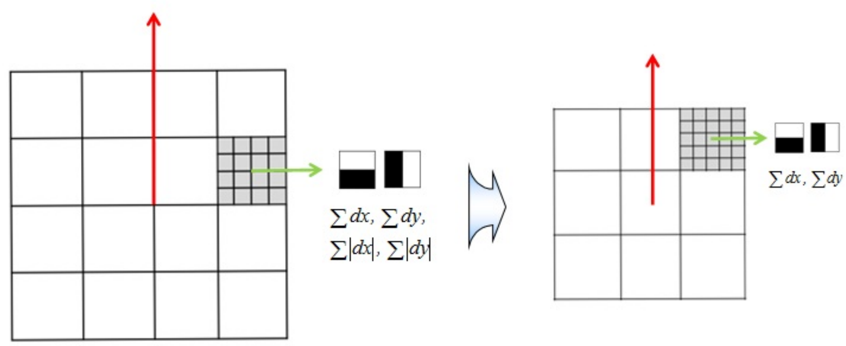


Figure 3: Reduction of dimension in feature descriptor.

Figure 3 describes the proposed method for decrement of complexity using reduction of the dimension of feature descriptor. The conventional algorithms using 64-dimension descriptor are not suitable for real time environment since their computational complexity for extracting the feature points is high. Therefore, the reduction of dimension in feature descriptor is necessary to effectively decrease computational complexity to carry out object recognition in real time environments [15]. The reduction of dimension in feature descriptor is used for calculated direction vector through scale s to determine dominant orientation and expanding its window to $\pi/2$ for estimation of accurate dominant orientation by a lot of directional information. The rectangular window is divided 3×3 sub-region and then sub-regions are re-divided into 5×5 sub-region. In Equation 3, 18($3\times3\times2$)-dimension feature descriptor in segmented regions makes up two feature vectors.

$$V_{sub} = [\Sigma dx,\ \Sigma dy] \tag{3}$$

The sum of the Haar wavelet which responses in horizontal (dx) and vertical (dy) directions are calculated. Since the Haar response is robust lighting condition, the proposed method decreases computation complexity and is also lighting invariant.

3.2 Object Tracking

Continuously adaptive MeanShift (CamShift) algorithm was derived from MeanShift algorithm for color based object tracking [21, 19]. CamShift method makes the outcome of the last frame as the initial value of the next frame for MeanShift algorithm, and carries out those steps in iterative [23, 24]. The process of the CamShift algorithm is depicted in the following.

input : image from video sequence

Convert color space;
while *End of frames* **do**
  Initialize the size and location of the search window;
  Find the center position of the search window;
  Calculate probability distribution for color in the search window;
  Call MeanShift, and Calculate the new size and position of the search window;
  Get next frame;
end

Algorithm 1: CamShift algorithm

When the interest object size is scaled, CamShift method can adaptively adjust the target region in order to continue the tracking. CamShift algorithm can set the calculation region of the probability distribution to the whole image and choose the initial location of the 2D mean shift search window. After choosing an initial location, the zeroth moment and first moment of x and y are calculated. The zeroth moment is expressed as Equation 1 and the first moment for x and y is expressed as Equations 4 and 5.

$$M_{00} = \sum_x \sum_y I(x,y) \tag{4}$$

$$\begin{aligned} M_{10} &= \sum_x \sum_y xI(x,y) \\ M_{01} &= \sum_x \sum_y yI(x,y) \end{aligned} \tag{5}$$

where $I(x,y)$ represents the back projected probability distribution at position (x,y) within the search window. x and y represent the range of the search window and they are slightly larger than the mean shift search window. After calculation of moment, the mean position of interest object (the centroid) is computed with the following Equation 6.

$$x_c = \frac{M_{10}}{M_{00}},\ y_c = \frac{M_{01}}{M_{00}} \tag{6}$$

where centroid value of x and y is denote x_c and y_c respectively. For the next video frame, the size of search window is recalculated as a color probability distribution function of the zeroth moment. The final step is the above steps repeated to converge (the change of centroid is smaller than preset threshold). After object tracking, the size and angle of the target in the image can calculate the first and second moment of distribution of intensity in the search window. By calculating, the second moment for x and y can be expressed as Equation 7.

$$\begin{aligned} M_{20} &= \sum_x \sum_y x^2 I(x,y) \\ M_{02} &= \sum_x \sum_y y^2 I(x,y) \\ M_{11} &= \sum_x \sum_y xyI(x,y) \end{aligned} \tag{7}$$

As the search window is recalculated we need to update the size of horizontal and vertical axis and the angle, which are the detected distribution of intensity.

4 Experiments and Results

We have experimented object recognition and tracking method under windows environments with the Core i3 CPU 3.30GHz and 4GB RAM memory using Visual Studio 2013, as shown in Figure 4. We have tested many video sequences. The computing task focuses on target modeling and matching, and concentrates on feature point detection and matching, and is in proportion to tracking window size.

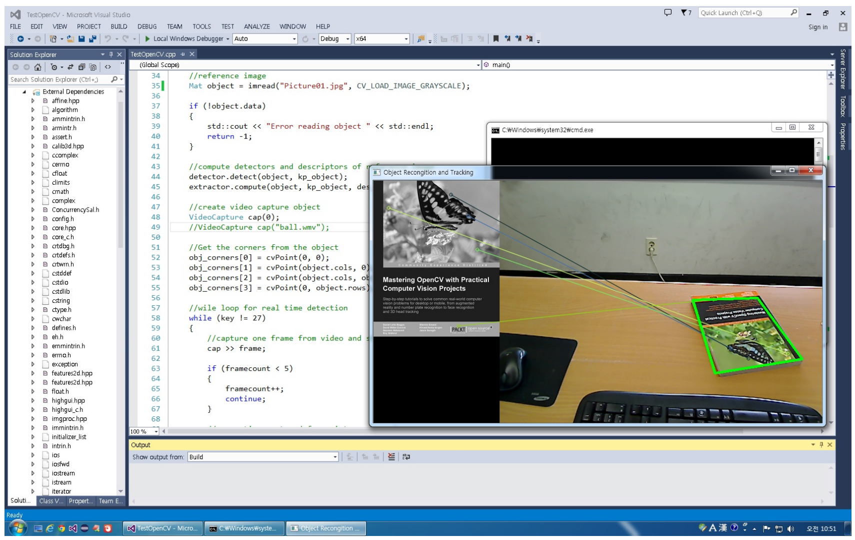


Figure 4: Implementation Environments.

SURF has low computational complexity for recognizing objects in real time. Therefore the proposed method extracts the features and finds out the corresponding matching points in a video sequence. Also we reduce computational complexity to efficiently decrease dimension in feature descriptor to carry out object recognition.

Figure 5, 6 and 7 describe object recognition and tracking result using this study's approach which is improved SURF and CamShift algorithm. And figure 8 shows that the interest target is accurately detected in complexity environment. Such is a low complexity and robust object recognition by advanced feature matching. Also it can carry out tracking efficiently. As the interest object changes, its region can adapt to resize the window correctly.

Table 1 shows improved performance in the comparison with existing algorithms of the proposed algorithm. The proposed method has efficiently decreased processing time to find the matching point through obtainment of the correct orientation information through an extended orientation window and reduction of the dimension of the feature descriptor. Therefore our method can improve problems, such as lost interest target when a similar color exists in the background, and high computational complexity to recognize an object using feature points.

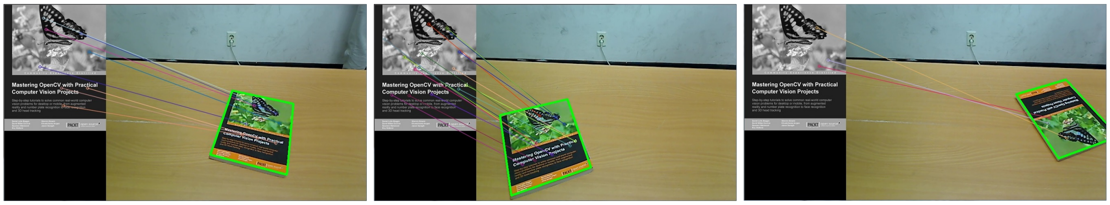


Figure 5: Object recognition and tracking results for book.

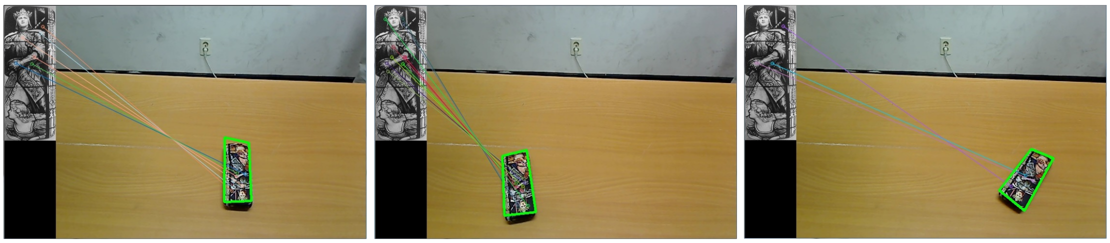


Figure 6: Object recognition and tracking results for case.

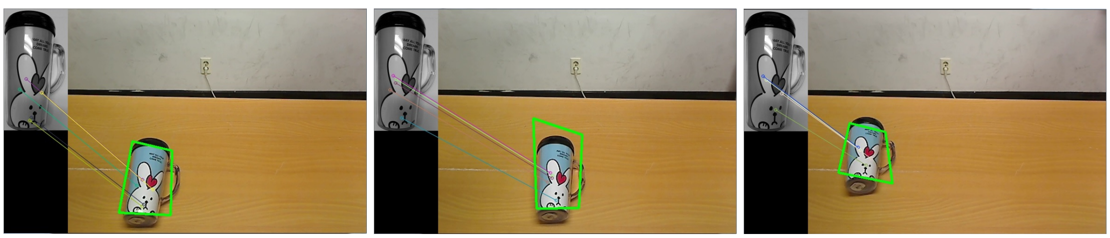


Figure 7: Object recognition and tracking results for cup.

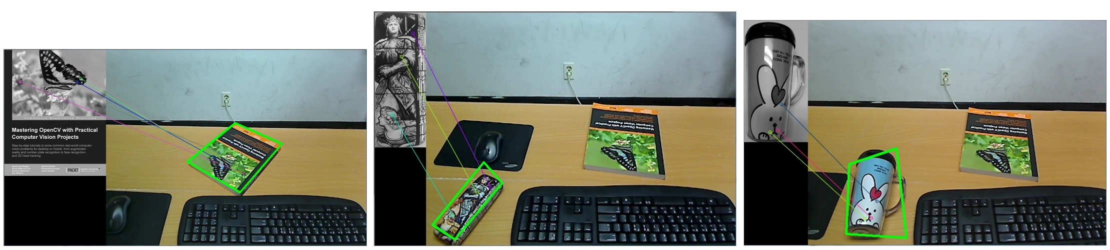


Figure 8: Object recognition and tracking result in complexity environment.

Table 1: Experimental Results.

Method	Recogniion Rate[%]	Recognition Time[sec]
Proposed Method	95%	0.49
SURF	94%	0.65
SIFT	96%	4.82

5 Conclusion

Recent researches on large-scale data processing have been actively carried out in the field of cloud computing. In video surveillance system, to efficiently track a moving object within a video sequence, we propose robust object recognition and tracking method, using advanced feature matching. The proposed algorithm recognizes objects with invariant features, and reduces dimensions of the feature descriptor to reduce computational time. The experimental result shows that our work is more fast and robust than the traditional methods and can track objects accurately in various environments. As future researches, our method needs to detect objects in a simple environment and recognize multiple objects on surveillance system.

Acknowledgments

This work was supported by Korea Sanhak Foundation for funding 2014.

References

[1] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346–359, 2008.

[2] M. M. Bhajibhakare and P. K. Deshmukh. Detection and tracking of moving object for surveillance system. *International Journal of Application of Innovation in Engineering and Management*, 2(12):298–301, 2013.

[3] M. Brown and D. Lowe. Invariant features from interest point groups. In *Proc. of the 2002 British Machine Vision Conference (BMVC'02), Cardiff, UK*, pages 656–665. British Machine Vision Association, September 2002.

[4] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5):564–577, 2003.

[5] S. A. Dave, M. Nagmode, and A. Jahagirdar. Statistical survey on object detection and tracking methodologies. *International Journal of Scientific and Engineering Research*, 4(3):1–8, 2013.

[6] M. Du, J. Wang, J. Li, H. Cao, G. Cui, J. Lv, and X. Chen. Robot robust object recognition based on fast surf feature matching. In *Proc. of the 2013 Chinese Automation Congress (CAC'13), Changsha, China*, pages 581–586. IEEE, November 2013.

[7] D. Exner, E. Bruns, D. Kurz, and A. Grundhofer. Fast and Robust CAMShift Tracking. In *Proc. of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'10), San Francisco, California, USA*, pages 9–16. IEEE, June 2010.

[8] A. R. Francois. Camshift tracker design experiments with intel opencv and sai. Technical Report IMSC-04-423, Institute for Robotics and Intelligent Systems, University of Southern California, August 2004.

[9] S.-W. Ha and Y.-H. Moon. Multiple object tracking using sift features and location matching. *International Journal of Smart Home*, 5(4):17–26, 2011.

[10] K. Huang, L. Wang, T. Tan, and S. Maybank. A real-time object detecting and tracking system for outdoor night surveillance. *Pattern Recognition*, 41(1):432–444, 2008.

[11] L. Juan and O. Gwun. A comparison of sift, pca-sift and surf. *International Journal of Image Processing*, 3(4):143–152, 2009.

[12] Y. Ke and R. Sukthankar. PCA-SIFT: A More Distinctive Representation for Local Image Descriptors. In *Proc. of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04), Washington, DC, USA*, volume 2, pages 506–513. IEEE, June-July 2004.

[13] Y.-H. Lee, J.-H. Park, and Y. Kim. Comparative analysis of the performance of sift and surf. *Journal of the Semiconductor & Display Technology*, 12(3):59–64, 2013.

[14] I. Leichter, M. Lindenbaum, and E. Rivlin. Mean shift tracking with multiple reference color histograms. *Computer Vision and Image Understanding*, 114(3):400–408, 2010.

[15] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[16] J. Pons, J. Prades-Nebot, A. Albiol, and J. Molina. Fast motion detection in compressed domain for video surveillance. *Electronics Letters*, 38(9):409–411, 2002.

[17] M. Stommel. Binarising sift-descriptors to reduce the curse of dimensionality in histogram-based object recognition. international journal of signal processing. *Image Processing and Pattern Recognition*, 3(1):25–36, 2010.

[18] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01), Kauai, Hawaii, USA*, volume 1, pages 511–518, December 2001.

[19] J. Wang, F. He, X. Zhang, and Y. Gao. Tracking objects through occlusions using improved kalman filter. In *Proc. of the 2nd International Conference on Advanced Computer Control (ICACC'10), Shenyang, China*, volume 5, pages 223–228. IEEE, March 2010.

[20] Q. Wang and Z. Gao. Study on a real-time image object tracking system. In *Proc. of the 2008 International Symposium on Computer Science and Computational Technology (ISCSCT'08), Shanghai, China*, volume 2, pages 788–791, December 2008.

[21] K. Werner and M. Kampel. Interest point based tracking. In *Proc. of the 20th International Conference on Pattern Recognition (ICPR'10), Istanbul, Turkey*, pages 3549–3552. IEEE, August 2010.

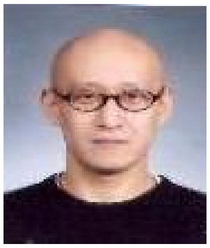
[22] A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. *ACM Computing Surveys*, 38(4):1–45, December 2006.

[23] Y. Yue, Y. Gao, and X. Zhang. An improved camshift algorithm based on dynamic background. In *Proc. of the 1st International Conference on Information Science and Engineering (ICISE'09), Nanjing, China*, pages 1141–1144. IEEE, December 2009.

[24] J. Y. Zhang, H. Y. Wu, S. Chen, and D. S. Xia. The target tracking method based on camshift algorithm combined with sift. *Advanced Materials Research*, 186:281–286, 2011.

[25] Z. Zhou, D. We, X. Peng, Z. Zhu, and K. Luo. Object tracking based on camshift with multi-feature fusion. *Journal of Software*, 9(1):147–153, 2014.

Author Biography



Yong-Hwan Lee received the MS degree in computer science and PhD in electronics and computer engineering from Dankook University, Korea, in 1995 and 2007, respectively. He was a Manager at KIS from 1995 to 2000, and at eKalos from 2000 to 2003. He developed ITA, EP, ERP solution with own BSA engine and language. He is working as member in an International Standard JPSearch, and he is still participated in SC29 WG1 JPEG AR. Currently, hhe is a Professor at the Department of Smart Mobile, Far East University, Korea. His research areas include image/video coding, image/video representation and retrieval, face recognition, augmented reality, mobile programming and multimedia communication.

Hyochang Ahn received the BS in computer science from Sangji University, Korea, in 2003 and received the MS degree and PhD in electronics and computer engineering from Dankook University, Korea, in 2006 and 2012, respectively. He was a Researcher at Jaein Industrial Technology, Korea, from 2009 to 2011. He was an Adjunct Professor at Far East University, Korea, from 2011 to 2012. Currently, he is working as Research Professor at Department of Applied Computer Engineering, Dankook University. His research interests include image processing, computer vision, embedded system and mobile programming.

Han-Jin Cho received the BS, MS and PhD in computer engineering from Hannam University, Korea in 1997, 1999 and 2002. Since then, he has been a Full Professor with the Department of Smart Mobile, Far East University, Korea. And he was awarded a certificate from the Ministry of Knowledge Economy, Korea in 2012. His main research interests include mobile applications and network security.

June-Hwan Lee received the BS, MS and PhD in Electrical Engineering from Dankook University, Korea in 1994, 1996 and 2001. Currently he is a Professor at the Department of Smart Mobile, Far East University, Korea. His research areas include Audio Processing, Multimedia Application and Smart Media.