

Multi-Target Tracking Using Color Information

Intaek Kim, Malik Muhammad Khan, Tayyab Wahab Awan, and Youngsung Soh

Abstract—For security purposes, it is prerequisite to track multiple targets efficiently. Most of the current implementation uses Kalman filter and color information independently. The proposed method combines extended Kalman filter and color information for tracking multiple objects under high occlusion. For tracking, the first thing done is the object detection. The background model used to segment foreground from background is spatio-temporal Gaussian mixture model (STGMM). Tracking consists of two steps: independent object tracking and occluded object tracking. For independent object tracking we exploit extended Kalman filter, whereas for occluded object tracking, color information attribute is used. The system was tested in real world application and successful results were obtained.

Index Terms—EKF, multi-target tracking, STGMM, tracking using color information.

I. INTRODUCTION

Tracking plays a vital role in the field of computer vision to interpret the behavior of people. Tracking people is used for many important applications like surveillance, intelligent control, automotive safety and virtual reality. These applications could only perform well if the detection of the people is quite precise. In order to track objects we need to extract objects first. Background subtraction is commonly used for detecting moving objects especially when background has not much changed. The most important issue in background subtraction is maintaining background. Detection of people in a scene is very hard task if two basic performance devaluating factors are present there: (a) illumination change, and (b) partial occlusion of interested objects by other objects.

Many background models have been proposed by researchers. Among them are running Gaussian average [1], Gaussian mixture model (GMM) [2], kernel density estimation [3], and eigenbackground [4]. An admirable review of these techniques is presented in [5], [6]. Difficulty of temporal models arises when there is a uniform motion in background such as camera jittering or non-uniform motion such as swaying tree branches, water rippling, and flag fluttering. To remedy, new background models were proposed that consider temporal behavior as well as spatial relations, so called spatio-temporal background models. In this work we use spatio-temporal Gaussian mixture model (STGMM) proposed by Soh *et al.* [7].

In this paper, we present a comprehensive approach in resolving these kinds of problems. We divide our work in four

important aspects. Firstly, we applied STGMM for object detection. Our proposed STGMM minimizes the noise in object detection and also excludes the shadow of the object. Secondly, we applied the extended Kalman Filter to predict the state of the object in the next frame, which enhances the object tracking and deals with the objects not being detected by STGMM in one or more frames. Thirdly, we used color information to track objects under high occlusion. The color attribute precisely helps to track the objects separated after being merged in STGMM. Lastly, we extracted the attributes of the objects i.e., its track, color and time of appearing and disappearing the scene. We stored this information in separate files for each tracked object. These attributes can be later used for searching a specific object in a surveillance video i.e. a car or a human with a certain color and specific range of time.

II. RELATED WORKS

Different methods have been proposed for multi-target tracking. However, in this section we describe several well-known techniques reviewed in three different domains: 1) GMM; 2) Tracking object using Kalman filter; and 3) Color histogram-based matching.

A. GMM

GMM proposed in [2] describe a background pixel using a mixture of K Gaussian distributions. So, it can deal with more complex background scenes, such as flapping flags and waving branches. The probability that the observed pixel is background is the weighted sum of the K Gaussian distributions. In order to avoid costly matrix computation, it is assumed that the R, G and B color channel have the same variance. Every time when a new pixel comes, it is checked with the already existing K distributions, until a match is found. If no match is found, then a new distribution is generated with the current pixel. After every updating process, the K distributions are reordered. So, the most likely background distribution is always at the top of the K distributions. Then the first X distributions are chosen which are above certain threshold, to represent the real background.

Zhen Tang and Zhenjiang Miao in [8] propose a new and fast background modeling method in comparison to [2]. It is assumed that each channel has its own variation. Secondly, to get faster adaption of the mean and variance values, the pdf is cut off from the learning rate. This proposed method computes three parameters less than [2]. Thus, performance is comparatively higher than the model proposed in [2].

B. Tracking with Kalman Filters

Kalman filter recursively estimates the state of target object; hence in tracking it is a useful technique which predicts the states of the moving objects. A recursive solution for linear

Manuscript received November 4, 2013; revised January 23, 2014.

The authors are with the Myongji University, Yongin, 449-728, Korea (e-mail: kit@mju.ac.kr, soh@mju.ac.kr, malikmk2003@yahoo.com, tayyabwahab90@gmail.com).

optimal filtering is proposed by R. E. Kalman [9] in 1960. Since then, an extensive research has been done in various domains such as in navigation systems. Thus, a wide range of literature is available on this subject. But, here we are limiting our review to Kalman filtering for multi-object tracking.

A broad exposure to Kalman filtering can be seen in [10]; whereas, a comprehensive review on Kalman filter in visual tracking is bring forth in [11]. Nguyen *et al.* [12] applied Kalman filter in distributed tracking system for tracking multiple moving people in a room using multiple cameras. Whereas Chang *et al.* [13] use both Bayesian network and Kalman filtering to solve the correspondence problem between multiple objects. In [14], a video surveillance system is proposed where detection, recognition and tracking of object is carried out. The performance of the approach is dependent on the proposed detection and recognition algorithms. Vector Kalman is proposed in [15] for tracking objects. In this paper, separate methods for occlusion and merge are applied to handle the confusing situations.

C. Color Histogram-Based Matching

Feature-based matching can help to solve the problems under confused situations; the objects are occluded and merged. The well describable features of an object are shape, color and temporal properties. Each feature has its contribution in distinguishing an object of a class from another but the color is the most useful feature which provides appropriate information to discriminate an object from another.

A common technique is color-based matching using RGB color histogram [16]. In this approach, color histogram technique is utilized to precisely match objects through indexing. This approach is extensively used for various tracking algorithms [17], [18]. However, the robustness is affected under the illumination change.

Color-based tracking methods can also be used along other techniques. Liu *et al.* [19] propose a tracking algorithm which utilizes both color histogram and particle filtering. Likewise, Limin [20] uses color-based Kalman particle filter object tracking algorithm. In crowded environments, a hue-saturation histogram with a particle filter based probabilistic technique is implemented in [21] for tracking objects.

Jia *et al.* [22] propose a technique for number plate matching of vehicles. In this proposed approach, the Gaussian weights are used to develop the corresponding relationship between the distance of two histograms and their Gaussian relationship. However, the scope of this technique is limited to the selected application. Also, because of the computational complexity, it can't be applied for tracking high resolution images and scenes which might contain multiple objects.

Correlation-Weighted Histogram Intersection (CWHI) technique proposed in [23] exploits Kalman filter with proposed CWHI for tracking multiple moving objects in confusing states i.e., inter-object occlusion and separation. Each moving object is assigned an individual Kalman tracker which is assisted by proposed manager (i.e., managing algorithm) during the entire process.

III. THE PROPOSED METHOD

The proposed method tracks multiple objects in a video applying extended Kalman filter and color information. As the color information is being used, the proposed method can even track efficiently under high occlusion. The proposed method consists of four steps. We are applying STGMM, extended Kalman filtering, extracting dominant color and storing the tracked information i.e., object class, object number, dominant color and time of entering and leaving the scene. Detailed description of these steps follows.

A. STGMM

In this step, we review the STGMM proposed by Soh *et al.* [7]. The notations used in [24] are followed. The density parameter set of k^{th} Gaussian distribution is supposed $N_k = \{m_k, s_k\}$ and the total parameter set as $Q = \{w_1, \dots, w_k, N_1, \dots, N_k\}$. Here the mean, standard deviation and priori weight are represented as m , s and w respectively. Then, for temporal GMM (STGMM), the posterior probability $P(k | X, Q)$ of sample X can be represented by (1).

$$P(k | X, Q) = P(k) f_{X|k}(X | k, N_k) / f_X(X | Q) \quad (1)$$

where, sum of the Gaussian mixture

$$f_X(X | Q) = \sum_{k=1}^K P(k) f_{X|k}(X | k, N_k) \quad (2)$$

and $P(k)$ is a priori probability of k^{th} Gaussian given by

$$f_{X|k}(X | k, N_k) = \frac{1}{\sqrt{2\pi}s_k} e^{-\frac{(X-m_k)^2}{2s_k^2}} \quad (3)$$

Model update is performed by (4) through (7).

$$\hat{m}_{k,t} = (1 - r_{k,t})m_{k,t} + r_{k,t}X_t \quad (4)$$

$$\hat{s}_{k,t}^2 = (1 - r_{k,t})s_{k,t}^2 + r_{k,t}((X_t - \hat{m}_{k,t})^T (X_t - \hat{m}_{k,t})) \quad (5)$$

$$r_{k,t} = a_t P(k | X_t, Q) / \hat{w}_{k,t} \quad (6)$$

$$\hat{w}_{k,t} = (1 - a_t)w_{k,t} + a_t P(k | X_t, Q) \quad (7)$$

where, for weight at time t , the learning ratio is a_t . Let $F(X) = \{F_1(X), F_2(X), \dots, F_n(X)\}$ denote n neighbors of pixel X including itself. For 3x3 neighbors $n = 9$. Let $\text{trans}(X)$ be the number of transitions in M previous classification results for pixel X . The transition count $\text{trans}(X)$ is increased by one if there is a transition from background to foreground or from foreground to background. Thus the maximum of transitions i.e., $\text{trans}(X)$ could be $M-1$. If the transition count is high, then the pixel under consideration is not of foreground. Such behavior can be seen due to swaying tree leaves and fluttering flags. So, such pixel will be called "UNSTABLE" and the pixels with low transition count will be called "STABLE".

Now we can describe the proposed STGMM algorithm as follows.

The STABILITY of each pixel could be determined using $M/3$ as a threshold. Thus if the transition count is greater than or equal to $M/3$, then the pixel is called “UNSTABLE” and otherwise “STABLE”.

B. Extended Kalman Filtering

For tracking, we adopt to use and follow the notations used for EKF in [25].

Extended Kalman Filter is preferred over linear Kalman filter when the state variables and measurements are no longer linear combination of state variables, inputs to the system and noise. For example, moving objects such as people or cars on the road are moving straight most of time, but people can stop or change the walking speed and car can make either right or left turn. In this case the linearity does not hold. So, extended Kalman filter is used to overcome such nonlinearity.

The key variables used in KF or EKF are state estimate (\hat{x}_k) and measurement (z_k) whose relation can be depicted in Fig. 1.

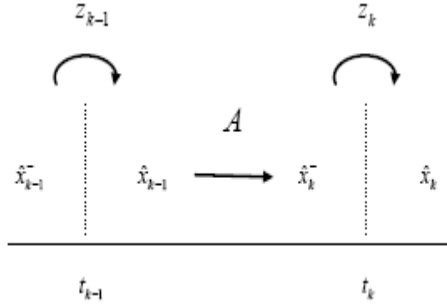


Fig. 1. Estimation and prediction in KF/EKF.

Kalman filter as well as extended Kalman filter find the estimate (\hat{x}_k) which is the best approximation of the variable x_k by iterating update and prediction. The estimate (\hat{x}_k) is updated or corrected by the prediction (\hat{x}_k^-) and the current measurement (z_k). This process is called update. On the other hand, prediction is to obtain predicted state (\hat{x}_k^-) from the previous estimate (\hat{x}_{k-1}).

The complete set of EKF equations is shown below in Fig. 2.

In Fig. 2 the external input is the measurement (z_k) and final output is state estimate (\hat{x}_k). Deriving the state estimate by prediction and update, several parameters (A_k , H_k , Q , R) are required for system model. A_k and H_k are $\frac{df}{dx}$ and $\frac{dh}{dx}$ respectively and they are extracted from the nonlinearity $\hat{x}_k = f(\hat{x}_{k-1})$ and $h(\hat{x}_k) = H_k \hat{x}_k$. Another system parameter Q is the covariance matrix of process, whose diagonal terms represent error square. Lastly R represents the characteristic of noise in the measurement. Like KF, EKF finds the estimate (\hat{x}_k) from the measurement (z_k). The prediction in Fig. 2 computes predicted value of estimate (\hat{x}_k^-) and predicted error covariance from which

Kalman gain and error covariance are obtained. In the update, Kalman gain is calculated using several parameters and state estimate is updated with measurement. In addition, error covariance is also updated.

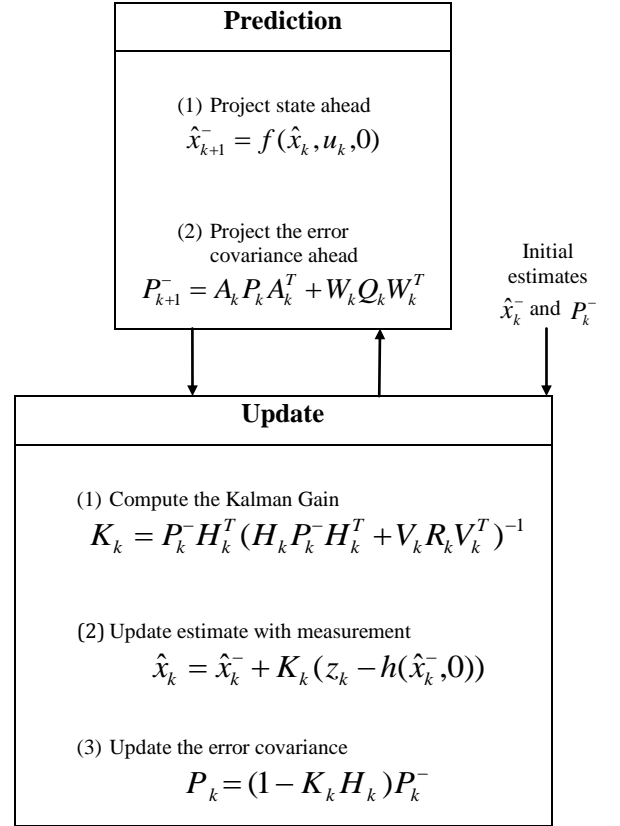


Fig. 2. A complete operation of EKF with the equations.

C. Extracting Dominant Color

When objects are occluded and shown as merged blobs in STGMM for several frames, they would be treated as new objects by the simple extended Kalman tracking algorithm. And extended Kalman filter will give them new tracking numbers. Hence, tracking of objects throughout the scene would not be possible with a single tracking number. The extended Kalman filter alone cannot overcome this problem. We need some other attributes which could help us in reassigning the object tracking number back after they separated from merged or occluded condition. There are some attributes that are maintained, termed as invariant attributes, and there are other attributes that are changed, termed as variant attributes.

Invariant attributes include object class and object color, whereas variant attributes include object size and object position. We use invariant attributes to track objects under high occlusion. When objects are tracked, we attempt to decide the kind of objects. Objects will be classified as vehicles, human, human group, and unidentified. Then color attribute is determined as follows:

The foreground extracted from STGMM is taken into consideration. The RGB values of each pixel are found out and categorized into n^3 classes, where n equals 4. For each blob the two most frequent colors are found and they are taken as invariant attribute of the object. Then the objects before merging are compared to the objects after the separation to reassign their objects numbers.

*Beginning of Algorithm**For each image**Apply STGMM using (1) and (7)**For each Foreground**Find First and Second Most Frequent Colors**Dominant Colors = Frequent Colors**End of For loop**End of For loop**For each object X**If New Dominant Colors (after separation) = Previous Dominant Colors (before merging)**Same object X**End If**Else**New object Y**End Else**End of For loop**End of Algorithm*

The algorithm given above clearly shows the two most occurring colors of the objects are extracted and then used for comparing, after they separate to reassign them their tracking number correctly. This algorithm is called immediately after the merged objects separate. This helps us to distinguish between new appearing objects and the objects merged and separated the merged objects separate. This helps us to distinguish between new appearing objects and the objects merged and separated.

IV. EXPERIMENTAL RESULTS

The experimental results are presented which shows the good tracking of moving objects under confusions. The results are presented in Fig. 3. The left half of the figure is the proposed extended Kalman filter with color information and right half represents the STGMM of each frame.

Fig. 3(a) left half shows two objects with different colors are entering the scene in frame number 1094. Their STGMM is shown on right half of the Fig. 3(a). They are given tracking number 6 and 7 and their dominant colors are stored as invariant attribute. In the 1123 frame these objects merged and shown as single blob in STGMM window. The Fig. 3(b) shows they are merged and our algorithm stores information of both these objects before merging and merged object numbers are shown in the output. When these objects separated in frame number 1156 in Fig. 3(c), the object number 6 gets tracking number back based on its dominant color. And the object in the middle has new dominant color so it is updated with new tracking number 8. When this blob number 8 breaks in STGMM, it breaks into two blobs and hence they are numbered 7 and 8. So, using the dominant color information we successfully reassigned tracking numbers to the respective objects. This is shown in Fig. 3(d) in frame 1209 of the test video.

The presented work is based on our initial research analysis on multi-object tracking using linear Kalman filter. The previous technique was good for separated objects but for confusing situation it failed. So, we figured out that dominant color could be used for matching the class of the objects.



(a) Objects detected



(b) Objects merged



(c) Object getting original tracking number after demerging



(d) All objects with their original tracking number after demerging

Fig. 3. Experimental results of EKF using color information.

V. DISCUSSION

Multi-target tracking is hard task when the objects are exhibiting confusing situation i.e., high occlusion. Color information based tracking helps to address this problem. The proposed method in paper not only tracks efficiently but also provides data files of the tracked objects for later review. This technique can be applied to crowded scenes where it is hard to track multiple objects and to find the desired object in couple of hours of surveillance videos seems impossible. Additionally, the proposed method can perform efficiently over poor background modeling.

The tracking is robust in terms of handling the conflicted situations due to occlusion but sensitive to detection technique. In future, we will also be investigating the behavior and event of the object associated to tracking. We would like to track multiple objects across the viewpoints using single dome camera based on Kalman prediction and color information.

ACKNOWLEDGMENT

This work (Grants No. C0005448) was supported by Business for Cooperative R&D between Industry, Academy,

and Research Institute funded by Korea Small and Medium Business Administration in 2012.

REFERENCES

- [1] C. Wren, A. Azarbayejani, T. Darrel, and A. Pentland, "Pfinder: real-time tracking of the human body," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 780-785, 1997.
- [2] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, 1999, pp. 246-252.
- [3] A. Elgammal, R. Duraiswami, D. Harwood, and L. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *Proceedings of IEEE*, vol. 90, no. 7, pp. 1151-1163, 2002.
- [4] N. Oliver, B. Rosario, and A. Pentland, "A Bayesian computer vision system for modeling human interactions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 831-843, 2000.
- [5] M. Piccardi, "Background subtraction techniques: a review," in *Proc. 2004 IEEE International Conference on Systems, Man, and Cybernetics*, 2004, pp. 3099-3104.
- [6] Y. Benezeth, P. Jodoin, B. Emile, H. Laurent, and C. Rosenberger, "Review and evaluation of commonly-implemented background subtraction algorithms," in *Proc. 19th International Conference on Pattern Recognition*, 2008, pp. 1-4.
- [7] Y. S. Soh, Y. S. Hae and I. T. Kim, "Spatio-temporal Gaussian mixture model for background modeling," in *Proc. 2012 IEEE International Symposium on Multimedia (ISM)*, Dec. 2012, pp. 360-363.
- [8] Z. Tang and Z. J. Miao, "Fast background subtraction and shadow elimination using improved Gaussian mixture model," presented at IEEE International Workshop on Haptic Audio Visual Environments and Their Applications, Oct. 2007.
- [9] R. E. Kalman, "A new approach to linear filtering and Prediction Problems," *Transactions of the ASME, Journal of Basic Engineering*, vol. 82, pp. 34-45, 1960.
- [10] P. S. Maybeck, *Stochastic Models Estimation and Control: Voloum 1*, New York: Academic Press, 1979.
- [11] N. Funk, "A study of the Kalman filter applied to visual tracking," University of Alberta, Project for CMPUT 652, 2003.
- [12] N. Nguyen, H. H. Bui, S. Venkatesh, and G. West, "Multiple camera coordination in a surveillance system," *ACTA Automatica Sinica*, vol. 29, pp. 408-422, 2003.
- [13] T. H. Chang and S. Gong., "Tracking multiple people with a multi-camera system," in *Proc. IEEE Workshop on Multi-Object Tracking*, 2001, pp. 19-26.
- [14] H. Yu, Y. Wang., F. Kuang, and Q. Wan, "Multi-moving targets detecting and tracking in a surveillance system," in *Proc. the 5th World Congress on Intelligent Control and Automation*, China, June, 2004, pp. 5253-5257.
- [15] S. A. Vigus, D. R. Bull, and C. N. Canagarajah, "Video object tracking using region split and merge and a Kalman filter tracking algorithm," in *Proc. ICIP*, 2001, pp. 650-653.
- [16] M. Swain and D. Ballard, "Color Indexing," In *Proc. the Third IEEE International Conference on Computer Vision*, Japan, 1990, pp. 11-32.
- [17] R. Collins, Y. Liu, and M. Leordeanu. "Online selection of discriminative tracking features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1631-1643, 2005.
- [18] T. Yang, S. Z. Li, Q. Pan, and J. Li., "Real-time multiple objects tracking with occlusion handling in dynamic scenes," in *Proc. the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, pp. 970-975.
- [19] F. Liu, Q. Liu, and H. Lu, "Robust color-based tracking," in *Proc. 3rd International Conference on Image and Graphics*, 2004, pp. 132-135.
- [20] X. Limin, "Object tracking using color-based Kalman particle filters," in *Proc. IEEE International Conference on Signal Processing*, 2004, pp. 679-682.
- [21] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet, "Color-based probabilistic tracking," in *Proc. 7th European Conf. on Computer Vision*, 2002, pp. 661-674.
- [22] W. Jia, H. Zhang, X. He, and Q. Wu., "Symmetric colour ratio gradients in spiral architecture," presented at the 18th International Conference on Pattern Recognition, 2006.
- [23] S. S. Pathan, A. Al-Hamadi, and B. Michaelis, "Intelligent feature-guided multi-object tracking using Kalman Filter," in *Proc. 2nd International Conference on Computer, Control and Communication*, 2009, pp. 1-6.
- [24] P. Power and J. Schooness, "Understanding background mixture models for foreground segmentation," in *Proc. Image and Vision Computing New Zealand*, pp. 267-271, 2002.
- [25] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 29, issue 3, pp. 401-422, 2004.



Intaek Kim was born in Seoul, Korea in 1960. He received BS and MS in electronics engineering from Seoul National University in Seoul, Korea in 1980 and 1984 respectively. He obtained PhD in electrical engineering from Georgia Institute of Technology in Atlanta, Georgia, USA in 1992.

He worked for Goldstar Central Research Lab from 1993 to 1995 as a senior engineer and joined Myongji University from 1995. He is now a professor in the

Dept. of Information and Communication Engineering. His recent publications deal with the area of face recognition, hypersepctral image and MR imaging.

His research interest includes pattern recognition, image processing and smart grid area.

Prof. Kim is a member of Korean Institute of Electronics Engineer.



Malik Muhammad Khan was born in Lahore, Pakistan in Nov. 8 1988. He got BS in electrical engineering in 2011 from Govt. College University in Lahore, Pakistan. He entered a master course in information and communication engineering in Myongji University in 2012.

His current interest of research includes vehicle to grid, Kalman filtering, background modeling and multi-viewpoint tracking.



Tayyab Wahab Awan was born in Peshawar, Pakistan in Aug. 17 1990. He got BS in telecommunication engineering in 2012 from National University of Computer and Emerging Sciences in Peshawar, Pakistan. He entered a master course in information and communication engineering in Myongji University in 2013.

His current interest of research includes vehicle to grid, Kalman filtering, particle filtering and multi-viewpoint tracking.



Youngsung Soh was born in Seoul, Korea in Mar. 4 1956. He got BS in electrical engineering in 1978 from Seoul National University in Seoul, Korea. He obtained MS and PhD in computer science from the University of South Carolina in Columbia, South Carolina, USA in 1986 and 1989, respectively.

He served in the Korean Army from June 1980 to Sept. 1982. He worked in Systems Engineering Research Institute in Korea as a senior researcher from Sept. 1989 to Feb. 1991. He joined Myongji University in Korea from Mar. 1991 and is currently a full professor in the Dept. of Information and Communication Engineering.

His current interest of research includes object tracking, stereo vision, and parallel algorithms for image processing. Prof. Soh is a member of Korea Information Processing Society and Korea Signal Processing Systems Society.