

Optimization and Tracking of Vehicle Stable Features Using Vision Sensor in Outdoor Scenario

Kajal Sharma

Flat No. 301, Building 206, Piorville Apartment, 24 Namyang-Dong, Seongsangu, Changwon, Korea
Email: kajal175@gmail.com

Abstract—Detection and tracking of stable features in moving real time video sequences is one of the challenging task in vision science. Vision sensors are gaining importance due to its advantage of providing much information as compared to recent sensors such as laser, infrared, etc. for the design of real-time applications. In this paper, a novel method is proposed to obtain the features in the moving vehicles in outdoor scenes and the proposed method can track the moving vehicles with improved matched features which are stable during the span of time. Various experiments are conducted and the results show that features classification rates are higher and the proposed technique is compared with recent methods which show better detection performance.

Index Terms—Vision Sensor, Image segmentation, Machine vision, Object recognition.

I. INTRODUCTION

Vision sensors such as cameras, stereo vision cameras, and infrared cameras have many advantages, including size, weight, and power consumption, over other sensors such as radar, sonar, and laser range finder, etc [1]. Vision sensors are also uniquely suited for tasks for target tracking and recognition and have the capability of enabling navigation in terrain environments [2]. Vehicle tracking via video processing is of great interest which can yield new parameters including lane change and vehicle trajectories along with vision information. The vehicle trajectories are measured over a length of roadway rather than at a single point and the vision data is used to provide additional information from the vehicle trajectories. For this, the most important requirement is the automatic segmentation of target vehicle from the background. The other requirements are traffic conditions and lighting conditions which includes traffic congestion, varying speeds in different lanes, sunny, overcast, night, rainy, etc [3].

Scale-invariant feature transform (SIFT) is given by Lowe to obtain the features invariant to image scaling and rotation, and partially invariant to changes in illumination and view. In addition, Lowe proposed an object recognition system with the use of distinctive SIFT

features for robust tracking for augmented reality with change in scene. The technique is based on nearest-neighbor search, but this method requires high computation time in the training phase of initialization. To overcome the limitation of the recent tracking algorithms, this paper proposed an optimized vehicle tracking approach to track the stable vehicle features efficiently in less time with the use of vision sensor. The proposed method tracks the stable and neuro optimized matched features in the segmented objects of the traffic sequence.

In this paper, a feature-based tracking scheme for fast vehicle detection is proposed. In the proposed scheme, all the moving vehicles are tracked with the fast feature matching between the different segmented vehicles, and the system performs the efficient vehicle tracking in outdoor environment by using segmented stable vehicle features which are optimized by the neural network methodology.

This paper is organized as follows: Section II present the related work for the vehicle tracking and feature based methods. Section III discusses the proposed method scheme and the experiments for tracking the moving vehicles are detailed. In Section IV, the tracking results are shown and the outcomes of the proposed method is discussed which shows advantages over the recent tracking techniques in terms of computation times. Section V concludes this paper.

II. RELATED WORKS

A variety of tracking algorithms is given in the literature for tracking the objects for e.g., global template based tracking [4-5] and feature based tracking algorithms [6-7] are given to track the objects with templates and features. Some of the tracking algorithms are based on manually-defined models and few others are based on training the targets during initialization [8]. In others methods, the objects are separated from the background to handle the difficulties of the change in appearance [9] which consists of a model with a discriminative classifier.

An algorithm has been proposed by Grabner et al. to obtain the features with online boosting method [10]. Numerous vehicle detection algorithms have been

proposed including background subtraction, frame differencing and motion based methods [11-12]. However, vehicle detection in the dynamic scenes remains a challenging task to detect and track the moving vehicles.

In [13], a local-feature point's configuration method is introduced for vehicle classification using computer graphics (CG) model images. This method used the eigen-window approach and has several advantages such as it detect the vehicles even if it changed its path due to veering out of the lanes and also if parts of the vehicles are occluded. This technique requires high computations for real time images for the target vehicles. In [14-17], other feature-based algorithms are given to gather the object features and to match the images of vehicle objects.

A subregion based novel statistical method is presented in [18] to detect the vehicles automatically based on local features of subregion. In this method, principal components analysis (PCA) weight vector is used to pattern the low-frequency components and an independent component analysis (ICA) coefficient vector is used to pattern the high-frequency components which is used for recognition non-occluded and partially occluded vehicles. A multimodal temporal panorama (MTP) method is given in [19] for vehicle detection and reconstruction. In this method a remote multimodal (audio/video) monitoring system is used to extract and reconstruct vehicles in real-time motion scenes.

Table 1. Review of vehicle tracking, feature based method and the observations for different conditions

Algorithm	Observation
Global template based tracking [3-4]	track the objects with templates and features, based on manually-defined models
Feature based tracking algorithms [5-6]	based on manually-defined models and few others are based on training the targets during initialization
Grabner feature tracker [9]	obtain the features with online boosting method
Novel statistical method [18]	automatic vehicle detection which depend on local features of subregion using PCA
Babenko online tracker [24]	tracker with multiple instances learning to track the objects with feature samples
Vehicle classification algorithm [13]	based on computer graphics (CG) model images to configure the local features
SIFT feature matching algorithm [14]	image matching and feature gathering

In addition to detection and motion estimations, multimodal approach helped in the reconstruction process of vehicles, resulting into dealing the occlusion, motion blurring and differences in perspective views situations. A new method is proposed in [20] to detect moving vehicles which used versatile movement histogram technique.

In [21], an iterative and distinguishable framework for vehicle classification is proposed in which framework is based on edge points and modified SIFT descriptors but has the limitation of image size and quality and large intra-class dissimilarities [22]. An object recognition system with the use of distinctive SIFT features for robust tracking is introduced in [23] for augmented reality with change in scene. In this, matching is performed with nearest-neighbor search which requires high computation in the offline-training phase.

Babenko et al. [24] introduced an online tracker with multiple instances learning to track the objects consisting of positive and negative bags of feature samples. In [25], a semi-supervised learning algorithm is given with online classifier with positive and negative samples based on structural constraints. A number feature based techniques for image matching and feature gathering is introduced by Lowe [14].

III. PROPOSED METHOD AND EXPERIMENTS

In the proposed method, the tracking of vehicles is accomplished by segmenting the vehicles and obtains invariant features. The vehicle segmentation varies with

change in shapes and colors, and obtains the segmented object by background subtraction. In this approach, subtracting the vehicle object from the input image is done in the segmentation process which highly reduces the search space in the tracking stage. However, detection of vehicles in moving videos using background subtraction techniques is not an easy task due to change in the outdoor scenarios such as change in movements of the background and light in the environment. It is still an unsolved problem to detect stable features continuously in the changing scenarios with the movement of the vehicles; the initial step in the proposed approach is the separation of the moving vehicles from the background with the use of motion information. The detailed block diagram of the proposed methodology is illustrated in Fig. 1 and matched features in the different motion vehicles is shown in Fig. 2 for the different time span.

A. Vehicle Feature Classification

In this approach, the vehicle feature classification in different vehicle dataset is accomplished with the SIFT descriptors which is rotation and scale invariant. Lowe proposed SIFT invariant feature descriptor to match features of objects in different objects under different invariant conditions [14]. In the proposed scheme, these descriptors dataset are followed by the classified feature reduction step to optimize the classified feature set. For each segmented vehicle object, an orientation histogram was formed based on local image gradient directions to assign an orientation to each feature keypoint location. Each vehicle object feature consists of 4×4 array of

histograms with 8 orientation bins and consists of 128 dimensional descriptor vectors. The matched features in the segmented object in different time span for one vehicle object is shown in Fig. 2. The clusters of motion vectors between different vehicle frames are obtained by further segmenting vehicles into smaller patches at first, and then assigning each patch to one of the clusters. Each segmented vehicle consists of an approximate hundreds of feature descriptors which consumes lots of memory space for each frame in the traffic video; thus added computational overhead for long traffic video. Thus, the proposed technique is used to optimize the high dimensional descriptor by using neural network based algorithm. The detailed proposed technique is given below.

SIFT descriptor is obtained for each vehicle object by the computation of gradient magnitude and orientation at each object point in a segmented region around a center point. Each object region is divided into $n \times n$ subregions and an orientation histogram is computed for each subregion by accumulating subregion samples with gradient magnitudes. SIFT feature vector is generated by the concatenation of histograms from all the subregions.

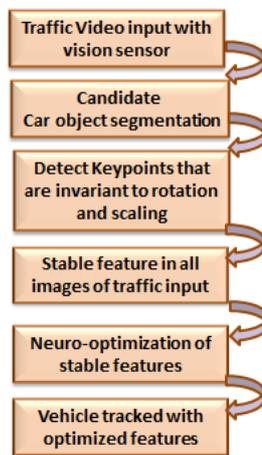


Fig.1. Block diagram of the proposed system.

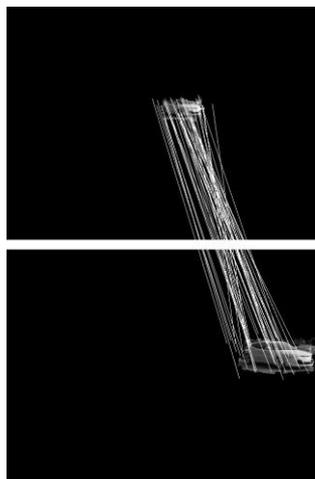


Fig.2. Matched features in the segmented vehicle in different time span; the above image shows two different video frames in which the vehicle position is changed with the change in time.

B. Feature-Based Vehicle Tracking

In the proposed technique, feature-based tracking is accomplished to track the similar feature points in the moving object with change in time at different time intervals. The similar feature points are computed in all the segmented vehicle objects in the different time instances to detect the invariant stable feature points. The features keypoints are extracted and clustered into features sets and perform matching in the similar features between different segmented objects in the traffic video. The vehicle objects are successfully recognized in the motion video by taking advantage of the repeatability of matched features which resulted into efficient tracking of moving vehicles.

For each segmented vehicle-object, coordinates and associated SIFT vectors for feature points in one segmented region define a feature. Let the number of points in segment i be J_i , the 2D coordinate of the j th point in segment i is $\vec{p}_{i,j}$, the SIFT vector of this point is $\vec{s}_{i,j}$. A feature f_i is a 3-tuple $\{\{\vec{p}_{i,j}\}, \{\vec{s}_{i,j}\}, \vec{c}_i\}$, $j = 1, \dots, J_i$, where $\{\vec{p}_{i,j}\}$ is the set of coordinates of all edge points in segment i , $\{\vec{s}_{i,j}\}$ is the set of SIFT vectors that are anchored at the edge points in segment i and \vec{c}_i is the average SIFT vector of segment i , i.e., mean of all $\vec{s}_{i,j}$ in segment i . Further, denote the set of all features of an object as $F = \{f_i\}$, $i = 1, \dots, N$, where N denotes the number of feature points of that region sample.

The descriptors are passed to the self-organizing map (SOM) network which is an unsupervised neural network. The dimension of the feature vectors in the feature pool are reduced by calculating the winner pixels. The vehicle objects feature vector is passed through the processing units in the input layer and the Euclidean distance between each feature vector and all of the weight vectors are computed for a predefined number of training cycles. The neuron with minimum Euclidean distance in the neighborhood region is marked as the best matching unit (BMU).

Let F_p denotes the pool of feature points. Randomly select a sample with all its features $F = \{f_i\}$, and store them into the feature pool $F_p = F = \{f_i\}$. Add another sample $F_{new} = \{F_{new}\}$ with all its features generated from the SIFT reduced descriptor set. For each F_{new} , compute the Euclidean-distance to all features in the feature pool. Suppose f_{min} in the feature pool has the smallest distance to F_{new} . If this smallest distance is less than a threshold, merge F_{new} with f_{min} in the feature pool by adding all its candidate points coordinates $\{x, y\}$ and SIFT vectors $\{F_{ij}\}$ to f_{min} and update the mean SIFT vector of f_{min} .

In order to match the segmented vehicle-object features in different images, the winner neuron pixel is calculated with the self-organizing map. The output neurons in the SOM are organized on a two-dimensional rectangular grid to reduce the descriptor feature dataset of the segmented vehicle. The input to the SOM consists of a set of vehicle-object features that are arranged in a two-

dimensional topological grid structure. The neurons in the SOM network are represented by a weight or prototype vector which consists of the number of components similar to the number of input feature sets, i.e. the dimensions of the input feature space. The SOM network is used to match the segmented vehicle-object features with the vehicle-object present in the next video frames as the position of the moving vehicle is changed in the next frames. The features of segmented vehicle-object are matched in the next frames by obtaining the winner pixel in the vehicle-object feature set.

IV. RESULTS AND DISCUSSION

For the performance evaluation of the proposed technique, test dataset of moving vehicle sequences were collected on the highly crowded roads in Changwon, South Korea. 320 by 240 image sizes were used for all the sequences with frame rate 30 fps. First the vehicles were segmented from the various backgrounds, and then with the proposed technique, the candidate target vehicle is successfully tracked by the stable optimized feature set. With the proposed method, the average vehicle detection rate was higher than 94%. Fig. 3 shows the results obtained by segmenting the vehicles for various datasets, i.e. truck, van, car. These are used for the tracking process.

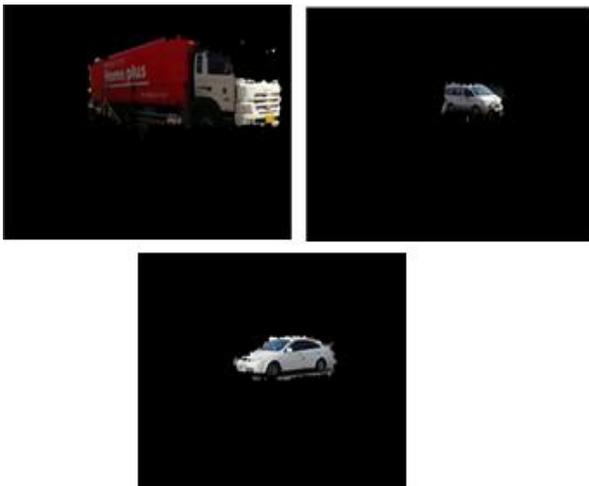


Fig.3. Segmented results for various vehicle dataset, i.e. truck, van and car.

Fig. 4 shows tracked vehicle results with the SIFT technique and the proposed technique for the Nth frame, N+10th frame, N+18th frame respectively. It can be observed that the tracking success rate of the proposed method is consistently higher than those of the Lowe's SIFT in different time instances. The features are neuro optimized with the feature reduction method and stable features are tracked continuously at different time interval. The average time to track the vehicle with the SIFT technique is 0.12578 sec which is reduced to 0.02528 sec with the proposed optimized method resulting into better performance in less time with high accuracy.



a. Tracked vehicle results with the SIFT technique and the proposed technique for the Nth frame.



b. Tracked vehicle results with the SIFT technique and the proposed technique for the N+10th frame.

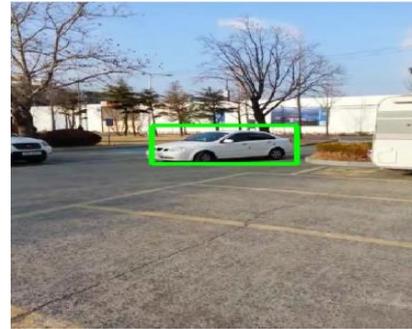




c. Tracked vehicle results with the SIFT technique and the proposed technique for the N+18th frame.

Fig.4. Tracked vehicle results with the SIFT technique and the proposed technique for the Nth frame, N+10th frame, N+18th frame in real-time video sequence.

The performance of the algorithm is examined by the analysis of the output results using various vehicle videos. Fig. 5 shows the results for the tracked white car with the SIFT technique and the proposed technique for the Nth frame, N+10th frame, N+18th frame respectively. The results show that the car is efficiently tracked at various time instances and the moved car can be located by the features at various viewpoints in less time as compared with the other state of art methods.

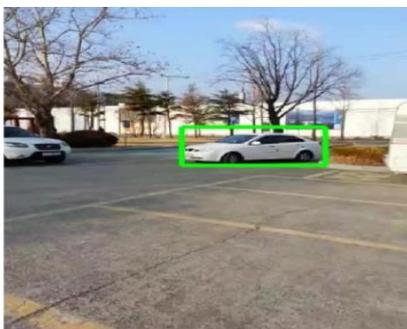


b. Tracked white car results with the SIFT technique and the proposed technique for the N+10th frame.



c. Tracked white car results with the SIFT technique and the proposed technique for the N+18th frame.

Fig.5. Tracked white car results with the SIFT technique and the proposed technique for the Nth frame, N+10th frame, N+18th frame in real-time video sequence.



a. Tracked white car results with the SIFT technique and the proposed technique for the Nth frame.



The proposed method attains promising improvement over the SIFT method in terms of computation time and accuracy. Fig. 6 shows the time-computation graph of the vehicle feature detection with the SIFT method, mean-shift, multiple instance learning algorithm and the proposed technique in different time spans of the moving vehicle. The system generates optimized stable features in different time intervals with accuracy of 96% accurate feature detection.

It is possible to obtain the less size descriptor in order to save memory space and can perform fast computation for stable features in matched segmented vehicles. The key modifications tuned to vehicle classification are that the down sampling of features is done in both the horizontal and vertical direction thus reducing the descriptor size by factor of two. Another modification is that during gradient orientation binning for histograms forming SIFT with $180 \pm$ differences are regarded as the same, which makes feature descriptor more robust with contrast differences and change in light.

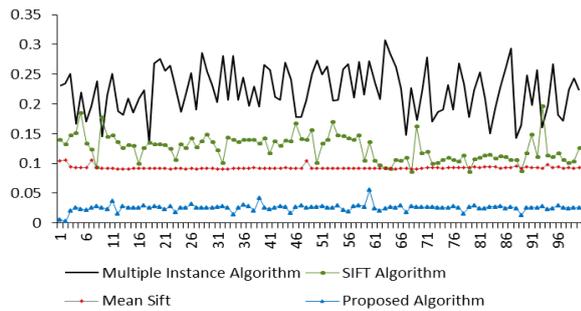


Fig.6. Computation time comparison with the SIFT technique, mean-shift, multiple instance learning algorithm and the proposed technique.

V. CONCLUSIONS

In summary, the system equipped with vision sensor is introduced for the fast and accurate vehicle tracking using the matched features in the different segmented objects under different time instances. The moving vehicles in real-time traffic videos were detected with the use of optimized matched features and the method includes an algorithm to detect and segment occlusive vehicles with neuro feature optimization.

In the experiment with feature classification, the proposed system was compared with the SIFT method, mean-shift, multiple instance learning algorithm. As a result, the proposed system successfully detects vehicle features and ensured stable tracking with matched features in less time. The proposed optimized method can be used for the development of real-time tracking and detection applications.

REFERENCES

- [1] R. Manduchi, A. Castano, A. Talukder, and L. Matthies, "Obstacle detection and terrain classification for autonomous off-road navigation," *Autonomous Robots*, vol. 18, no. 1, pp. 81–102, Jan. 2005.
- [2] E. Zamora and W. Yu, "Recent advances on simultaneous localization and mapping for mobile robots," *IETE Technical Review*, vol. 30, no. 6, pp. 490–496, Nov.-Dec. 2013.
- [3] Gargi, S. Dahiya, "A Gaussian Filter based SVM Approach for Vehicle Class Identification", *IJMECS*, vol.7, no.12, pp.9-16, 2015.
- [4] S. Avidan, "Support vector tracking," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, no. 8, pp. 1064–1072, Aug. 2004.
- [5] O. Williams, A. Blake, and R. Cipolla, "Sparse bayesian learning for efficient visual tracking," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1292–1304, Aug. 2005.
- [6] V. Lepetit, P. Laguerre, and P. Fua, "Randomized trees for real-time keypoint recognition," in *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pp. 775–781, June 2005.
- [7] M. Ozuysal, V. Lepetit, F. Fleuret, and P. Fua, "Feature harvesting for tracking-by-detection," in *Proc. 9th European Conf. on Computer Vision*, Graz, Austria, pp. 592–605, May 2006.
- [8] M. Isard, and J. MacCormick, "Bramble: a bayesian multiple-blob tracker," in *Proc. IEEE Int. Conf. on Computer Vision*, Vancouver, BC, pp. 34–41, July 2001.
- [9] X. Liu, and T. Yu, "Gradient feature selection for online boosting," in *Proc. IEEE Int. Conf. on Computer Vision*, Rio de Janeiro, pp. 1–8, Oct. 2007.
- [10] H. Grabner, and H. Bischof, "On-line boosting and vision," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 260–267, June 2006.
- [11] G. Monteiro, J. Marcos, M. Ribeiro, and J. Batista, "Robust segmentation for outdoor traffic surveillance," in *Proc. 15th IEEE Int. Conf. on Image Processing*, pp. 2652–2655, Oct. 2008.
- [12] W. Wang, J. Yang, and W. Gao, "Modeling background and segmenting moving objects from compressed video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 5, pp. 670–681, May 2008.
- [13] T. Yoshida, S. Mohottala, M. Kagesawa, and K. Ikeuchi, "Vehicle classification system with local-feature based algorithm using CG model images," *IEICE Trans. on Information and Systems*, vol.E85--D, no. 11, pp. 1745–1752, 2002.
- [14] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. of Computer Vision*, vol. 60, no. 2, pp. 91–110, Jan. 2004.
- [15] K. Mikolajczyk, and C. Schmid, "Scale & affine invariant interest point detectors," *Int. J. of Computer Vision*, vol. 60, no.1, pp. 63–86, Oct. 2004.
- [16] H. Bay, T. Tuytelaars, and L. V. Goo, "SURF: speeded up robust features," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, June 2008.
- [17] M. A. Akinlar, M. Kurulay, A. Secer, and M. Celenk, "A novel matching of MR images using gabor wavelets," *IETE Technical Review*, vol. 30, no. 2, pp. 129–133, Jan.-Feb. 2013.
- [18] W. Chi-Chen Raxle, and J. J. J. Lien, "Automatic vehicle detection using local features—A statistical approach," *IEEE Trans. on Intelligent Transportation Systems*, vol. 9, no. 1, pp. 83–96, March 2008.
- [19] W. Tao, and Z. Zhigang, "Real time moving vehicle detection and reconstruction for improving classification," in *Proc. IEEE Workshop on Applications of Computer Vision*, Breckenridge, CO, pp. 497–502, Jan. 2012.
- [20] W. Zhang, Q.M. Jonathan Wu, and H. Bing Yin, "Moving vehicles detection based on adaptive motion histogram," *Digit. Signal Process.*, vol. 20, no. 3, pp. 793–805, May 2010.
- [21] M. Xiaoxu, and W. E. L. Grimson, "Edge-based rich representation for vehicle classification," in *Proc. Tenth IEEE Int. Conf. on Computer Vision*, Beijing, pp. 1185–1192, Oct. 2005.
- [22] I. Gordon, and D. G. Lowe, "Scene modelling, recognition and tracking with invariant image features," in *Proc. Int. Symposium on Mixed and Augmented Reality*, Arlington, VA, pp. 110–119, Nov. 2004.
- [23] B. Babenko, M. H. Yang, and S. Belongie, "Visual tracking with online multiple instance learning," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Miami, FL, pp. 983–990, June 2009.
- [24] Z. Kalal, J. Matas, and K. Mikolajczyk, "P-N learning: Bootstrapping binary classifiers by structural constraints," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, San Francisco, CA, pp. 49–56, June 2010.
- [25] T. Yoshida, S. Mohottala, M. Kagesawa, and K. Ikeuchi, "Vehicle classification system with local-feature based algorithm using CG model images," *IEICE Trans. on Information and Systems*, vol.E85--D, no. 11, pp. 1745–1752, 2002.

Authors' Profiles

Kajal Sharma received a B.E. degree in computer engineering from University of Rajasthan, India in 2005, and M.Tech. and Ph.D. degrees in computer science from Banasthali University, Rajasthan, India in 2007 and 2010. From October 2010 to September 2011, she worked as a postdoctoral researcher at Kongju

National University, Korea. From October 2011 to April 2013, she worked as a post-doctoral researcher at the School of Computer Engineering, Chosun University, Gwangju, Korea. Presently she is working as an independent researcher in Korea. She has published many research papers in various national and international journals and conferences. Her research interest includes Image and video processing, Neural networks, Computer vision, and Robotics.