Journal of Korea Multimedia Society Vol. 11, No. 6, JUNE 2008(pp. 860-868)

Human Detection in Overhead View and Near-Field View Scene

Sunghoon Jung[†], Byunghee Jung[††], Minhwan Kim[†††]

ABSTRACT

Human detection techniques in outdoor scenes have been studied for a long time to watch suspicious movements or to keep someone from danger. However there are few methods of human detection in overhead or near-field view scenes, while lots of human detection methods in far-field view scenes have been developed. In this paper, a set of five features useful for human detection in overhead view scenes and another set of four useful features in near-field view scenes are suggested. Eight feature-candidates are first extracted by analyzing geometrically varying characteristics of moving objects in samples of video sequences. Then highly contributed features for each view scene to classifying human from other moving objects are selected among them by using a neural network learning technique. Through experiments with hundreds of moving objects, we found that each set of features is very useful for human detection and classification accuracy for overhead view and near-field view scenes was over 90%. The suggested sets of features can be used effectively in a PTZ camera based surveillance system where both the overhead and near-field view scenes appear.

Key words: Human Detection, Overhead View Scene, Near-Field View Scene, Feature Extraction, PTZ Camera-Based Surveillance System

1. INTRODUCTION

In general, surveillance systems based on CCTV cameras can be used for not only security purposes like invasion prevention or burglarproof but also managing purposes like safety management for prevention of accidents, law observance, or penalty imposition. We know that human detection techniques play an important role in surveillance systems. So many research results [1-6] for detecting humans have been developed, which could be used in outdoor or indoor scenes. However most of human detection methods in outdoor scenes have been designed to be applicable to only the far-field view scenes. Therefore these methods may not be adopted in surveillance systems at dangerous area such as building construction workspace or crane working space at harbors, where overhead and/or near-field view scenes appear frequently.

On the one hand, PTZ (pan-tilt-zoom) cameras are widely used in surveillance systems these days because wide area can be covered with a few PTZ cameras. When the PTZ cameras are adopted in monitoring or watching wide area, overhead and/or near-field view scenes can appear frequently. However there are few research results that are applicable to human detection in these view scenes.

In this paper, a set of four features useful for human detection in overhead view scene and an-

※ Corresponding Author: Minhwan Kim, Address: (609-735) Department of Computer Engineering, Pusan National University, Busan, Korea, TEL: +82-51-510-2423, FAX: +82-51-517-2431, E-mail: mhkim@pnu.ac.kr

Receipt date : Oct. 31, 2007, Approval date : Apr. 10, 2008

[†] Dept. of Computer Engineering, Pusan National University
(E-mail: shjung@pnu.ac.kr)

[††] Dept. of Computer Engineering, Pusan National University
(E-mail: id_jbh@pnu.ac.kr)

[†††] Dept. of Computer Engineering, Pusan National University

※ This work was supported for two years by Pusan National University Research Grant

other one in near-field view scene are suggested. Those features are most useful ones in discriminating humans from other moving objects in each view scene and they are selected among feature-candidates that are related to geometrically varying characteristics of moving objects in video sequences.

2. CLASSIFICATION OF CAMERA VIEWS

In this paper, a camera view is classified into three types, far-field, near-field, or overhead view, according to the angle between the camera optical axis and the horizontal plane passing through the projection reference point of the camera and parallel to the ground, as shown Fig. 1. Fig. 2 shows three typical examples for each type of view.

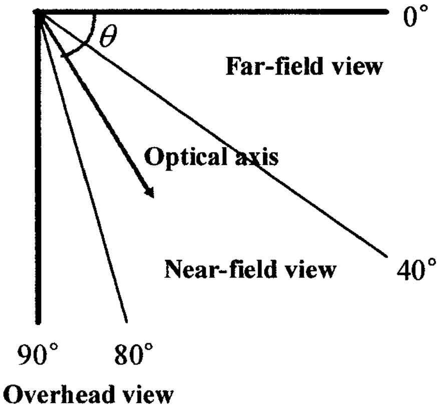


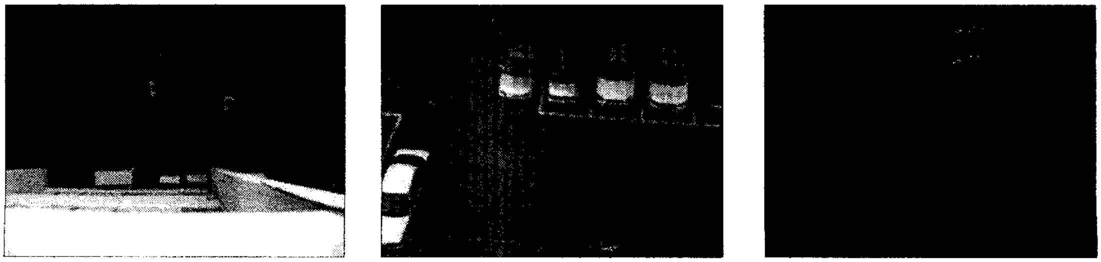
Fig.1. Classification of camera views

3. FEATURE-CANDIDATE EXTRACTION

Geometrically varying characteristics of moving objects in samples of video sequences are analyzed to extract feature-candidates that will be used in selecting a set of features for near-field view scene and another one for overhead view scene.

3.1 Moving object extraction and shadow removal

Moving regions are first extracted by using the difference image technique that computes the difference between the background image with no moving objects and the current image with moving objects. Then the moving regions are rectified by removing small noise regions and shadow regions that make hard to classify objects. There are many researches for shadow removal [7-10]. We used the shadow removal method [7,8] which removes shadows appropriately by utilizing the characteristic that shadows make intensity dark in some area and saturation low without large change on color information. Fig. 3 shows an example of moving object extraction with shadow removal and without shadow removal. In Fig. 3 (b) and (c), the shadow region is extracted and makes hard to analyze the geometrical characteristics of the moving object. After the shadow removal method [7,8] is applied to, we can extract only the moving object region as shown Fig. 3 (e) and (f). The rectified regions are defined as moving objects and classi-



(a) Overhead view (b) Near-field view (c) Far-field view

Fig. 2. Typical examples for each type view

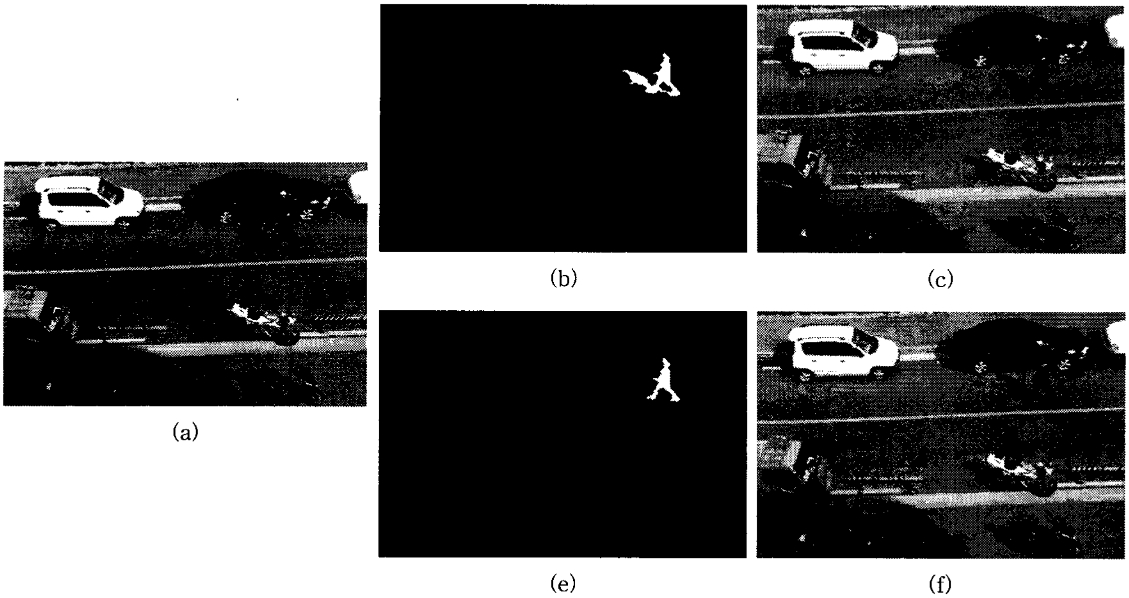


Fig. 3. An example of moving object extraction without and with shadow removal

fied into four object classes, human, human group, vehicle, and animal(dog).

3.2 Extraction of feature-candidates

In previous research results [1-6] for detecting humans in far-field view scenes, widely used classification features are compactness change-rate, motion direction change-rate, ratio of major and minor axis of minimum bounding box, area of an object, and dispersedness (= 1 / compactness). However all the features cannot be useful for detecting humans in overhead and/or near-field view scenes, because geometrical characteristics in these view scenes are different from ones in far-field view scenes. For example, only the human head and shoulders are shown in overhead view scenes, while whole body of human appears usually in far-field view scenes. So we looked out geometrical characteristics in overhead and near-field view scenes and found the following useful observations.

The orientation is the angle between least inertia axis of an object and x-axis of image. The compactness represents degree of circularity that is defined as 4π×area / (perimeter)^2. The elongation is ratio of major and minor axis of an object. The inverse MBB occupancy is ratio of the area of minimum bounding box (MBB) and the object area.

Fig. 4 shows variation of the geometrical features over whole frame sequences in overhead view scenes. The area feature is not tested, because its usefulness is dependent on applications

Geometrical features	Rigid objects	Non-rigid objects
Orientation change	very low	medium - high
Compactness (change)	high (low)	medium - high (high)
Elongation (change)	high (low)	medium - high (high)
Inverse MBB occupancy (change)	low (low)	low - high (low - high)
Area	medium - high	low - medium

even though it is very useful for itself. We see in Fig. 4 that the orientation and the compactness features of human and human group classes show very high variation because of their movements of arms and legs. On the other hand, those of vehicle (car) class show very low variation because it is a kind of rigid body. We can also see that mean value of each geometrical feature can be a good candidate feature in classifying moving objects into the four object classes. Fig. 5 shows variation of the selected geometrical features in near-field view scenes. We see that variation of orientation is very different from that in overhead view scenes.

Through general analysis of variations of the geometrical features in Fig. 4 and Fig. 5, we can infer that some measures related to variation over several frames are more useful for classifying moving objects than ones derived from only one frame. Thus mean and variance of each geometrical feature over 10 frames are selected as the

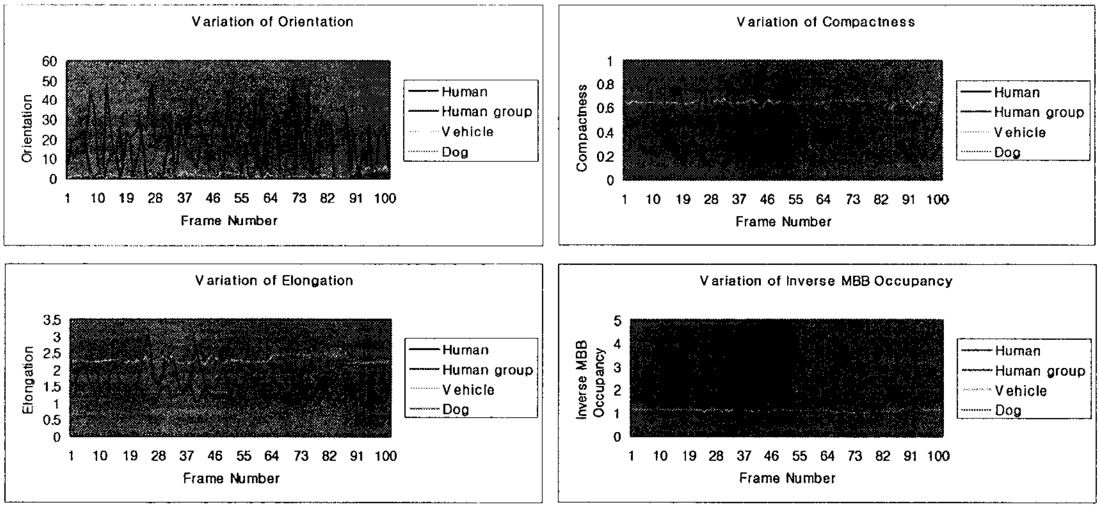


Fig. 4. Variations of geometrical features of moving objects in overhead view scenes

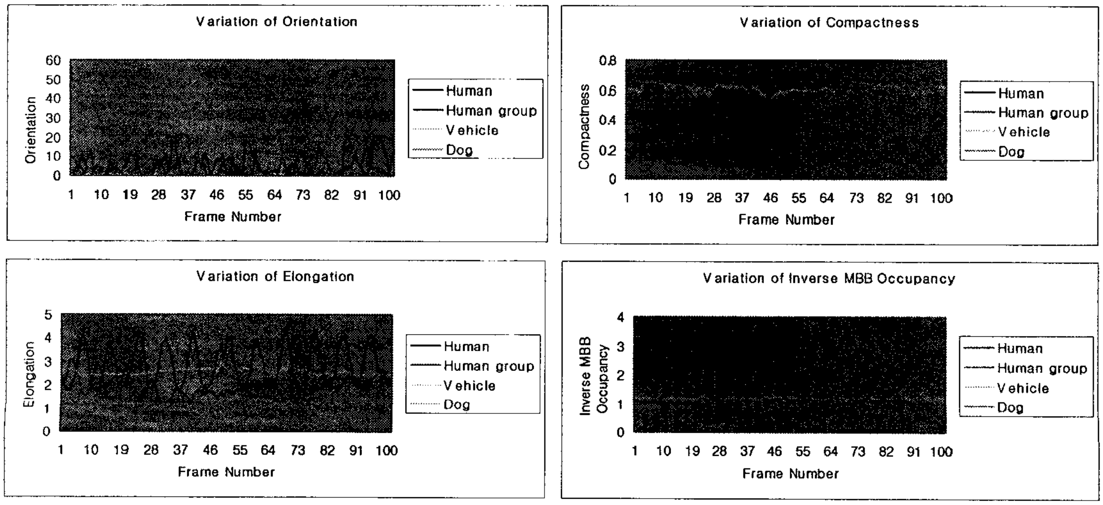


Fig. 5. Variations of geometrical features of moving objects in near-field view scenes

feature-candidates.

4. SELECTION OF USEFUL CLASSIFICATION FEATURES

Usefulness of the eight feature-candidates is tested with 14 video sequences of overhead view and 13 ones of near-field view by using the data mining tool, Weka3.5 [11]. The Weka uses the method called K-fold cross validation which divides samples into K sub-samples and uses one sample for test and other samples for training. After K times tests with different test sample at every time, the K test results are made into one by averaging. In general, the number of folds depends on the size of data set and we used 10 folds in our test. This method costs more computational time than simple test and training set method but provides more reliable results.

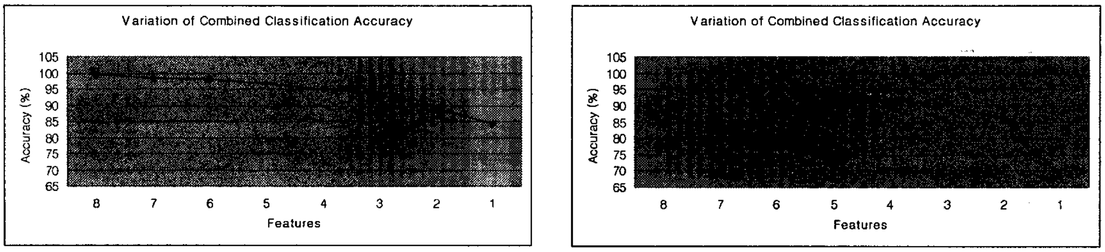
In our case, the video sequences are first split off video fragments with 10 frames long and then they are used for testing usefulness of each feature-candidate by using the back propagation algorithm [12] in Weka3.5. Through the usefulness tests, the feature-candidates are sorted in decreasing order based on the classification accuracy of each one. Table 1 shows the ordered list of the feature-candidates by their classification accuracy.

If we use only one feature, the classification accuracy is not much high. However, by combining the features, the classification accuracy can be improved much. Naturally the more features are used, the higher classification accuracy will be get. But one of the problems with many features is that not all the features are important for classification. This problem is connected with dimension reduction. By reducing the dimension of certain problem, we can save computational cost while maintaining the performance. Even though there were many researches for dimension reduction techniques [13], we tested every case of combination of feature-candidates because the number of feature-candidates in our case, eight, is not so large. Fig. 6 shows variation of the best combined classification accuracy with n feature-candidates.

The best combined classification accuracy falls as the number of feature-candidates is decreased as shown in Fig. 6, but there is no break point. Thus it is difficult to conclude how many feature-candidates are appropriate for classification of objects into the four object classes, human, human group, vehicle, and animal(dog). So we tried to observe contribution of each feature-candidate to the classification. We counted the feature-candidates

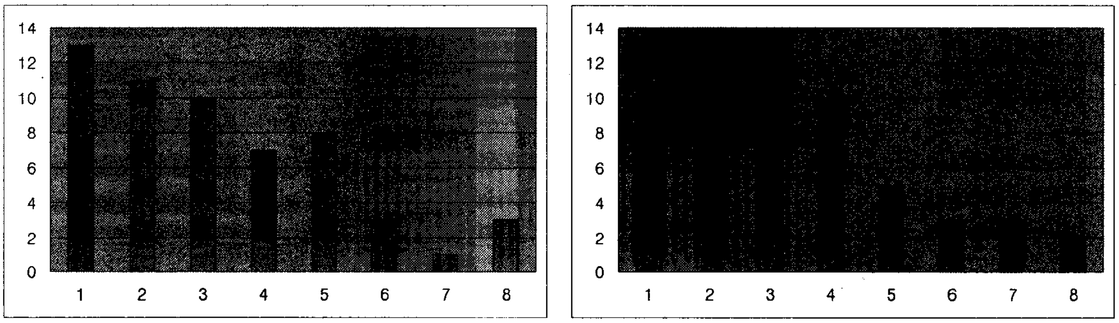
Table 1. Ordered list of the feature-candidates by classification accuracy

Overhead view	1. Variance of orientation	84.18%
	2. Mean of inverse MBB occupancy	82.76%
	3. Mean of orientation	73.57%
	4. Mean of compactness	65.10%
	5. Variance of inverse MBB occupancy	65.10%
	6. Mean of elongation	65.05%
	7. Variance of elongation	62.21%
	8. Variance of compactness	59.61%
Near-field view	1. Variance of elongation	68.35%
	2. Mean of elongation	66.98%
	3. Mean of compactness	66.69%
	4. Mean of orientation	56.09%
	5. Mean of inverse MBB occupancy	54.93%
	6. Variance of orientation	54.86%
	7. Variance of compactness	51.08%
	8. Variance of inverse MBB occupancy	49.56%



(a) (b)

Fig. 6. Variation of the best combined classification accuracy in the test with (a) video fragments of overhead view and (b) those of near-field view



(a) (b)

Fig. 7. Contribution of each feature-candidate to high accuracy of classification in (a) overhead view and (b) near-field view

included in the best two results of classification accuracy among the combined *n* feature-candidates tests (n = 1~7), thereby the counted number of each feature-candidate represents how many times it contributes to high accuracy of classification. Fig. 7 shows the contribution of each feature-candidate in overhead and near-field view scenes, respectively. The abscissa represents the ordered number of the feature-candidates in Table 1 and the ordinate shows the counted number.

We can see in Fig. 7 that the contribution of a feature-candidate tends to be high as its classification accuracy alone is high. On the one hand, there are some distinguishable feature-candidates in Fig. (a) and (b), which show relatively high contribution. In this paper, we selected the five topmost-contributed candidate-features in Fig. 7(a) as the useful features for classification of objects in the overhead view, while the four ones in Fig. 7(b) for classification of objects in the near-field view. Table 2 shows those useful features and it also gives some useful features [1-6]

Table 2. Useful feature sets for classifying moving objects in overhead, near-field, and far-field view scenes

Overhead view	Near-field view	Far-field view
Variance of orientation Mean of inverse MBB occupancy Mean of orientation Mean of compactness Variance of inverse MBB occupancy	Variance of elongation Mean of elongation Mean of compactness Mean of orientation	Compactness change-rate Motion direction change-rate Elongation Area Dispersedness

having been used for detecting humans in far-field view scenes. We see that useful features are quite different from each other.

5. EXPERIMENTAL RESULTS AND DISCUSSIONS

We made an experiment on 2,112 video fragments of overhead view and 1,378 ones of near-field view with the useful feature sets for overhead and near-field view by using the data mining tool, Weka3.5 [11].

Table 3 shows the classification results. The combined classification accuracies for overhead and near-field view are 96.59% and 92.52%, respectively. Table 4 gives the detailed report of misclassification. We see that there is relatively high confusion between human class and human group class. If an application does not need to distinguish human from human group and vice versa, the human detection accuracy reaches 98.86% for overhead view scenes and 97.67% for near-field view scenes. Therefore we conclude that the suggested sets of useful features are very useful for detecting humans.

6. CONCLUSIONS

A set of useful features for human detection in overhead view scenes and another one in near-field view scenes are suggested in this paper. Usefulness of the two sets was verified through experiments with several kinds of video sequences. We also found that the useful features in each view

Table 3. Classification results with various types of objects in overhead and near-field view scenes

	In overhead view scenes	In near-field view scenes
Extracted moving objects	Human: 539 Human group: 1074 Vehicle: 472 Dog: 27 Total: 2112	Human: 541 Human group: 471 Vehicle: 259 Dog: 107 Total: 1378
Accurately classified objects	2040	1275
Misclassified objects	72	103
Classification accuracy	96.59%	92.52%

Table 4. Detailed report of misclassification

	Real object types		Misclassified results	
Overhead view	Human	32	Human group	32
	Human group	22	Human Vehicle Dog	16 5 1
	Dog	18	Human Human group Vehicle	1 8 9
Near-field view	Human	17	Human group Vehicle	12 5
	Human group	68	Human Vehicle	59 9
	Vehicle	17	Human group	17
	Dog	1	Vehicle	1

were much different from the widely used features in far-field view. The suggested sets of useful features are expected to be effectively used in a PTZ camera based surveillance system where both the overhead and near-field view scenes appear.

REFERENCES

[1] P. Viola, M.J. Jones, and D. Snow, "Detecting Pedestrians Using Patterns of Motion and Appearance," *IEEE International Conference on Computer Vision*, pp. 734-741, Oct. 2003.

[2] R.T. Collins, et al., "A System for Video Surveillance and Monitoring," *CMU-RI-TR-0-12, VSAM Final Report*, Carnegie Mellon University, 2000.

[3] A.J. Lipton, H. Fujiyoshi, and R.S. Patil, "Moving Target Classification and Tracking from Real-Time Video," *Proc. Fourth IEEE Workshop on Applications of Computer Vision*, pp. 8-14, Oct. 1998.

[4] J. Zhou and J. Hoang, "Real Time Robust Human Detection and Tracking System," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol.3, pp. 149-156, June 2005.

[5] B. Bose and E. Grimson, "Learning to Use Scene Context for Object Classification in Surveillance," *Proc. Joint IEEE International Workshop on VS-PETS*, Nice, France, pp. 94-101, Oct. 2003.

[6] E. Rivlin, M. Rudzsky, R. Goldenberg, U. Bobomolov, and S. Lepchev, "A Real-Time System for Classification of Moving Objects," *Proc. 16th International Conference on Pattern Recognition*, Vol.3, pp. 688-691, Aug. 2002.

[7] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, and S. Sirotti, "Improving Shadow Suppression in Moving Object Detection with HSV Color Information," *Proc. IEEE International Conference Intelligent Transportation Systems*, pp. 334-339, Aug. 2001.

[8] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.25, pp. 1337-1342, Oct. 2003.

[9] A. Prati, I. Mikic, C. Grana, and M.M. Trivedi, "Shadow detection algorithms for traffic flow analysis: a comparative study," *Proc. IEEE International Conference Intelligent Transportation Systems*, pp. 340-345, Aug. 2001.

[10] J. Stander, R. Mech, and J. Ostermann, "Detection of moving cast shadows for object segmentation," *IEEE Transactions on Multimedia*, Vol.1, pp. 65-76, Mar. 1999.

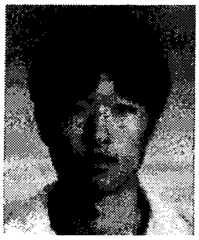
[11] http://www.cs.waikato.ac.nz/ml/weka/

[12] Y. LeCun, L. Bottou, G.B. Orr, and K.R. Muller, "Efficient Backprop in Neural Networks - Tricks of the Trade," *Springer Lecture Notes in Computer Sciences 1524*, pp. 5-50, 1998.

[13] I.K. Fodor, "A survey of dimension reduction techniques," *UCRL-ID-148494, LLNL technical report*, June 2002.

Sunghoon Jung

He received his B.S. and M.S. degrees from Pusan National University, Busan, Korea, in 2006 and 2008, respectively. He is currently a Ph.D. degree student of the Dept. of Computer Engineering in Pusan National University, Korea. His research interests include intelligent surveillance system and computer vision.



Byunghee Jung

He received his B.S. degree from Silla University, Busan, Korea, in 2007. He is currently a M.S. degree student of the Dept. of Computer Engineering in Pusan National University, Korea. His research interests include intelligent surveillance system and computer vision.

Minhwan Kim

He received his B.S., M.S., and Ph.D. degrees from Seoul National University, Seoul, Korea, in 1980, 1983, and 1988, respectively. He is currently a professor of the Dept. of Computer Engineering in Pusan National University, Korea. His research interests include intelligent surveillance system, multimedia information retrieval, color engineering, and computer vision.