# Example based Learning for Object Detection in Images

Taewan Kim
Pohang University of Science and Technology
Pohang, South Korea
taey16@postech.ac.kr

Daijin Kim
Pohang University of Science and Technology
Pohang, South Korea
dkim@postech.ac.kr

## ABSTRACT

In this paper, we describe a general learning architecture for object detection especially car detection. In order to build such a system, we first perform dimension reduction for each example by using maximizing mutual information criterion. The algorithm directly selects projection basis from examples which can minimize Bayes error. This algorithm is named as Maximizing Mutual Information(MMI) method. Given projection basis, all of examples are projected onto these basis and then trained by Support Vector Machine(SVM). This approach can be applied to any object with distinguishable patterns. In test process, we find objects in a image by using our exhaustive search algorithm which is called a Scale based Classifier Activation Map(SCAM). We applied our detection scheme into UIUC car/non-car database[2]. In this experiment we detect 181 cars in 170 images with 200 cars. This result is competitive comparing with other papers[1, 12].

## Categories and Subject Descriptors

I.5 [**Pattern Recognition**]: Application

## General Terms

algorithm, performance, experimentation

## Keywords

object detection, maximizing mutual information, dimensionality reduction, exhaustive search

## 1. INTRODUCTION

In surveillance system or mobile robot application, basically it is crucial to detect interesting or suspicious objects correctly. The main process of object detection system partitioned into both *feature extraction* and *searching objects*. The feature extraction process can be further categorized into a sort of methods based on the representation. Intensity based method is conventionally used for detect object such as human face[11, 13] or eigenvector and corresponding coefficient by using Principle Component Analysis(PCA) is used for object model[5]. Gradient also commonly applied for representing objects. [6, 17] use Histogram of Oriented Gradients(HoG) as their feature descriptor and then train a classifier to detect objects. Shape based representation that uses Active Shape Model(ASM) is also proposed by [14]. In this paper, mutual information based feature extractor is applied for feature descriptor as if [5] make use of PCA as their feature extractor. In general, mutual information based descriptor shows more discriminative power than PCA[12, 15]. Searching objects in a image is another main process in order to detect object. This process inherently rely on feature extractor and classifier so that there is no generally applicable methods. We therefore introduce an efficient detection scheme named as Scaled based Classifier Activation Map(SCAM). The idea conceptually based on ensanble classifier which means aggregation of classification result corresponding weak classifiers gives more discriminative power comparing with the result of each classifier. The remainder of this paper is organized as follows. Overview of the approach is mentioned in section 2. Details of dimensionality reduction methods and how to tune important parameters in Support Vector Machine(SVM) training process are explained in section 3 and section 4, respectively. In section 5, we propose our exhaustive search algorithm and show that how to works. Section 6 is devoted to presenting the experimental result of our car detection system as well as the experimental configuration. Finally, we conclude our approach and leave aspects of improvement in future work in section 7.

## 2. OVERVIEW OF THE APPROACH

Our example based learning approach can be divided into 4 processes; Each process is explained briefly as follows:

1. *Preprocessing* : Training examples need to perform illumination gradient correction and histogram equalization so that they can reduce lighting components, heavy shadow and compensate for differences in illumination, brightness, and camera gamma curve, respectively[11]. It is used for mining pure pattern reducing noise and illuminance.

2. *Feature Extraction* : A original training example in general represented by tremendously high dimensional vector so that it may contain noisy components which have a bad influence on classification task as well as it requires to spend large amount of time on processing or

*VNBA'08*, October 31, 2008, Vancouver, British Columbia, Canada.
Copyright 2008 ACM 978-1-60558-313-6/08/10 ...$5.00.

even it is impossible to handle. We therefore perform dimensionality reduction using a criterion which is to maximize mutual information between class label and feature vector.[12] Because this criterion assures minimizing Bayes error, the algorithm reduces dimension of each training example with high separability.

3. *Training a classifier* : Well trained classifier influences the classification result directly. It is mandatory that we tune all parameters related to the classifier. In our case, we make use of Support Vector Machine(SVM) and thus we should determine kernel type, kernel option and slack variable as well as the optimal number of dimension. There is no analytic way to determine such parameters so that we tune these parameters in empirical manner.

4. *Searching object in a image using trained classifier* : After training classifier, we proposed a detecting scheme which is called Scale based Classifier Activation Map (SCAM). In this scheme, it is commonly acceptable that although searching window is not exactly matched to the object, evaluated value by the trained classifier is amount of positive value in case that there are an object in arbitrary position at the window. In each iteration, we apply scaled window(actually fixed size of window but scaled image) to find objects in whole image. If the object is in the window per each scale, then the evaluated value is to be positive and this value is cumulated to the map at corresponding position and size of the window. After searching the object with variously scaled window, we normalize this map whose peak value is to be one. And then by thresholding this map in a certain value, we can detect the object.

## 3. FEATURE EXTRACTION

Extracting feature efficiently is very important problem in computer vision as well as machine learning and pattern recognition area. If we use whole image vectors in training process, there may be noisy features that disturb reducing Bayes error rate. We therefore need to reduce input dimension especially considering discriminative property. However in practice reducing too many dimensions can return undesirable result. Reducing too less dimension also can give a problem because we in general have limited number of samples and thus it occurs "*curse of dimensionality*". Therefore, we should find optimal number of dimension empirically by using efficient feature extraction method. The method should be able to transform original data into compact and discriminative feature. In order to meet both requirements, we make use of relative entropy between joint probability of class label and feature value, and product of both marginal probabilities with respect to these two random variables. It is known as mutual information. The theoretical background is explained in following section.

### 3.1 Mutual information

Suppose continuous random variable $\mathbf{y}_i \in \mathbf{Y}$, where $\mathbf{Y} \in \mathbb{R}^d$, and class labels $c_i \in \mathbf{C}$, where $\mathbf{C} = \{1, 2, \ldots, N_c\}$, $i = [0, N]$. $N_c$, and $N$ is the number of classes and samples, respectively. In Shannon's definition, information is referred to as uncertainty and its expected value is known as entropy

$$H(c) = -\sum_{c \in \mathbf{C}} p(c) \log p(c) \tag{1}$$

Assume that we draw one sample in Y at random. And then the uncertainty in terms of class prior probability is formulated as (1). If we observe continuous random variable $\mathbf{y}$, then this entropy is to be conditional entropy

$$\begin{aligned} H(c|\mathbf{y}) = & \; -\int_{\mathbf{y}\in\mathbf{Y}} p(\mathbf{y}) \Big( \sum_{c\in\mathbf{C}} p(c|\mathbf{y}) \log p(c|\mathbf{y}) \Big) d\mathbf{y} \\ = & \; -\sum_{c\in\mathbf{C}} \sum_{\mathbf{y}\in\mathbf{Y}} p(c,\mathbf{y}) \log p(c|\mathbf{y}) \qquad (2) \end{aligned}$$

The mutual information $I(C, Y)$ is defined as

$$\begin{aligned} = & \; \sum_c \int_{\mathbf{y}} p(c,\mathbf{y}) \log \frac{p(c,\mathbf{y})}{p(c)p(\mathbf{y})} d\mathbf{y} \qquad (3) \\ = & \; -\sum_c \int_{\mathbf{y}} p(c,\mathbf{y}) \log p(c) d\mathbf{y} - \big( -\sum_c \int_{\mathbf{y}} p(c,\mathbf{y}) \log p(c|\mathbf{y}) d\mathbf{y} \big) \\ = & \; H(C) - H(C|Y) \qquad (4) \end{aligned}$$

It easily can be derived such that (4), via simple Bayes rule. It means the amount of entropy in terms of class label reduced by conditional entropy is referred to as mutual information. (3) is analogous to Kullback-Leibler divergence so that it can be interpreted as distance measure between joint distribution with respect to $c$, $\mathbf{y}$ and product of each marginal distribution of these two random variables. It is noticeable fact that if the amount of (3) becomes close to one, the random variables $c$ and $\mathbf{y}$ become more statistically dependent each other. This phenomenon therefore quite intuitively say that if we can maximize mutual information, then these two random variable $c$ and $\mathbf{y}$ can be statistically related to each other which we desire. One thing remarkable is that if we only want to maximize(or minimize) $I(C, Y)$, it is alternatively sufficient to minimize(or maximize) $H(C|Y)$ stated in (2), since the expected uncertainty of class label, $H(C)$ is the same which is independent of $\mathbf{y}$.

### 3.2 Relation between mutual information and Bayes error rate

Relation between mutual information and Bayes error is given by Fano[7] and Hellman,Raviv[8] such that

$$\frac{H(C) - I(C,Y) - 1}{\log m} \leq P_e(Y) \leq \frac{1}{2}\Big(H(C) - I(C,Y)\Big), \tag{5}$$

where $P_e(Y)$ is classification error. And $m$ is the number of classes. These Bayes lower and upper error bounds commonly represent the fact that maximizing mutual information with respect to class label $C$ and feature $Y$ can optimize Bayes error, directly. It is key role in justifying the usage of $I(C, Y)$ as a proxy to classification error.

### 3.3 Dimensionality Reduction

In this section, we explain the method of computing transformation function which performs dimensionality reduction with compaction and discrimination by using a criterion that maximizes mutual information between class label and feature value. There are two major algorithms conducting dimension reduction by using information theoretic criterion[12, 15]. One is to select transform basis maximizing mutual information from training examples directly[12]. The other is to find rotation parameter with respect to all examples using gradient ascent method[15]. These two algorithms commonly use information theoretic criterion, however, their approach is quite different. In this paper, we use

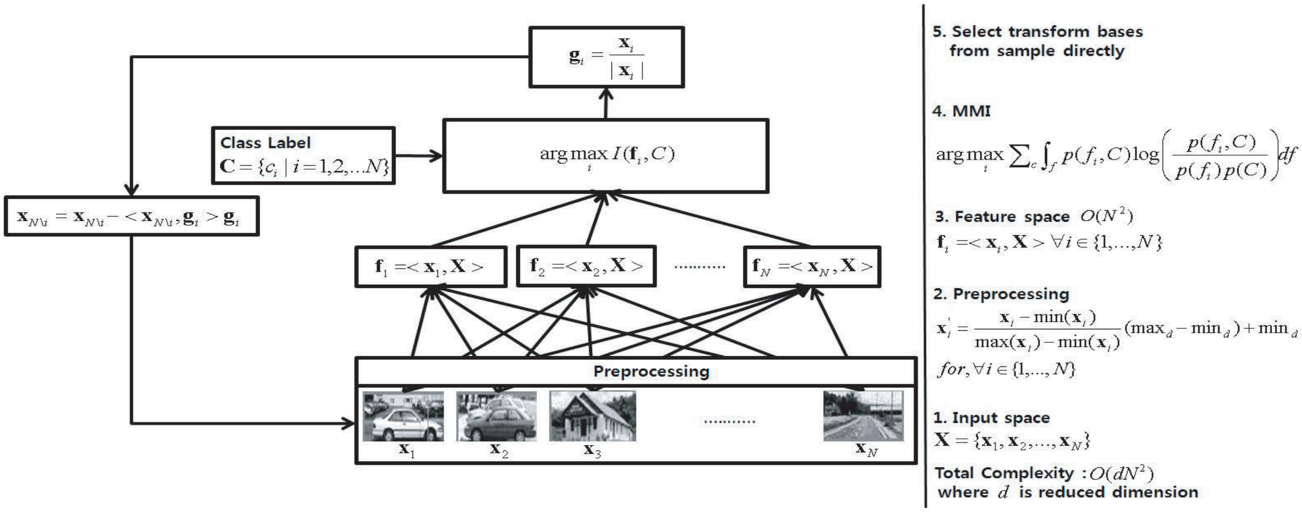
$\mathbf{g}_i = \frac{\mathbf{x}_i}{|\mathbf{x}_i|}$
Class Label $\mathbf{C} = \{c_i \mid i = 1,2,\ldots N\}$
$\arg\max_i I(\mathbf{f}_i, C)$
$\mathbf{x}_{N\backslash i} = \mathbf{x}_{N\backslash i} - <\mathbf{x}_{N\backslash i}, \mathbf{g}_i > \mathbf{g}_i$
$\mathbf{f}_1 =< \mathbf{x}_1, \mathbf{X} >$
$\mathbf{f}_2 =< \mathbf{x}_2, \mathbf{X} >$
..........
$\mathbf{f}_N =< \mathbf{x}_N, \mathbf{X} >$
Preprocessing
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
...........
$\mathbf{x}_N$
5. Select transform bases from sample directly
4. MMI
$\arg\max_i \sum_c \int_f p(f_i, C) \log\left(\frac{p(f_i, C)}{p(f_i)p(C)}\right) df$
3. Feature space $O(N^2)$
$\mathbf{f}_i =< \mathbf{x}_i, \mathbf{X} > \forall i \in \{1,\ldots,N\}$
2. Preprocessing
$\mathbf{x}_i' = \frac{\mathbf{x}_i - \min(\mathbf{x}_i)}{\max(\mathbf{x}_i) - \min(\mathbf{x}_i)}(\max_d - \min_d) + \min_d$
$for, \forall i \in \{1,\ldots,N\}$
1. Input space
$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$
Total Complexity : $O(dN^2)$
where $d$ is reduced dimension

**Figure 1: An illustration of Maximizing Mutual Information(MMI) method**

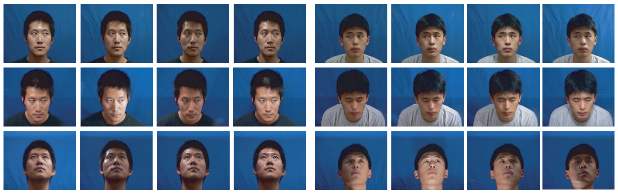

**Figure 2: Examples of training database which represent two people's face**

the former which is depicted in figure 1. First of all, we perform min-max normalization for all training examples in which values are ranged from 0 to 1. The training examples then are transformed into a feature space such that

$$\mathbf{f}_i^T = \mathbf{x}_i^T \cdot \mathbf{X}, \text{ for all } i \in \{1, \ldots, N\}, \tag{6}$$

where $\mathbf{x}_i$ is $i$th example, $\mathbf{X}$ is total examples in $\mathbb{R}^{d_{in} \times N}$, $d_{in}$ is a dimension of original training example and $N$ is total number of training images. We compute joint probability between this transformed feature value $\mathbf{f}_i$ and class label $c$. And then both distributions $p(\mathbf{f}_i)$ and $p(c)$ can be computed straightforwardly by marginalizing this joint probability with respect to these two random variables. These three distributions, $p(\mathbf{f}_i, c)$, $p(c)$ and, $p(\mathbf{f}_i)$, are used for computing mutual information formulated in (3). Among this $I(f_i, c)$ for all $i \in \{1, \ldots, N\}$, we select $i$th example which is the maximum value of $I(f_i, c)$. This $i$th example is the most suitable for minimizing Bayes error. The reason that we select $i$th example is already justified in (4). This selected example of course should be normalized such that $\mathbf{g}_i = \mathbf{x}_i / ||\mathbf{x}_i||$. And then all remaining examples should be orthogonalized with respect to the selected basis $\mathbf{g}_i$ such that

$$\mathbf{x}_{N\backslash i} = \mathbf{x}_{N\backslash i} - \left\{\mathbf{x}_{N\backslash i}^T \mathbf{g}_i\right\} \mathbf{g}_i \tag{7}$$

(6) can be think of as approximation to the orthogonalization between selected basis and reminder of total examples, $x_{N\backslash i}$. This process may be fused by random projection.[3, 10]. It means all selected basis $\mathbf{g}_i$ may be orthogonalized by the Gram-Schmidt process. All of the consecutive processes are iterated until we get $d$ number of basis. The $d$ is the number of reduced dimension such that $d_{in} >> d$. As we know in figure 1, total complexity of the Maximizing Mutual Information(MMI) method is $O(dn^2)$ because in order to transform original data into a feature space, we needs $O(n^2)$ times and this process is repeated up to $d$ times. In figure 3, we demonstrate the effectiveness of maximizing mutual information criterion. In this demonstration, we trained 600 images which represent two people's face that is shown in figure 2. It contains various pose and illuminance. The original training examples in $\mathbb{R}^{64 \times 48}$ space are transformed into $\mathbb{R}^2$ feature space that is plotted in figure 3. In figure 3, we can easily recognize that dimensionality reduction using maximizing mutual information outperforms comparing with Principle Component Analysis(PCA). In this case, the first and second basis generated by MMI which represent each of classes are more desirable than the first and second basis of PCA. It is quite nature because

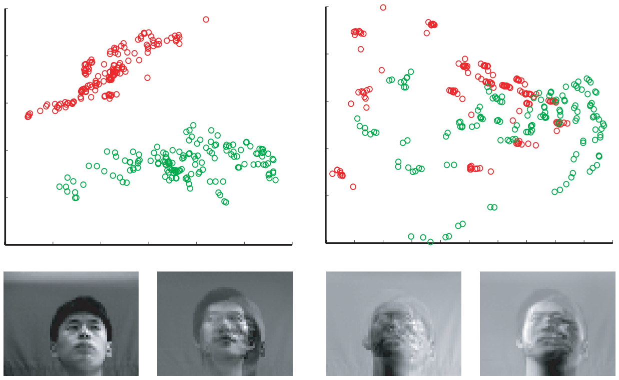

**Figure 3: Demonstration of dimensionality reduction using Maximizing Mutual Information(MMI) and PCA. Left-top: Dimension reduction using MMI. Right-top: Dimension reduction using PCA. Left-bottom: first and second basis computed by using MMI. Right-bottom: first and second basis computed by using PCA**

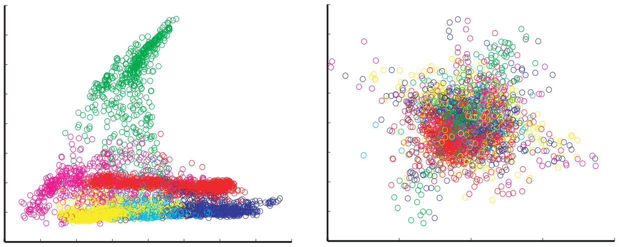

**Figure 4: Demonstration of dimensionality reduction Left: Dim. reduction using MMI. Right: Dim. reduction using PCA.**

**Table 1: Classification result(%) as changing the number of reduced dimension**

| Dimension | 1 | 2 | 4 | 8 | 16 | 32 | 64 |
|---|---|---|---|---|---|---|---|
| PCA+SVM | 57.0 | 65.2 | 86.0 | 88.0 | 93.0 | 94.0 | 80.4 |
| MMI+SVM | 80.2 | 86.2 | 88.8 | 91.0 | 93.8 | 97.0 | 67.6 |

Principle Component Analysis(PCA) only considers maximizing variance of training data. However, Maximizing Mutual Information(MMI) considers not only compaction but also minimizing classification error. In figure 4, we also apply MMI method to the Landsat image database in UCI Machine Learning Repository[16]. The result shows the fact that MMI method is more effective than PCA as well.

## 4. TRAINING A CLASSIFIER

When training a classifier, we should tune the optimal number of dimension and certain parameters used for classifier. In our system, we use Support Vector Machine(SVM) implemented by [4] as our classifier. In such a case, we should tune parameters such as kernel type, corresponding kernel variables and slack variable as well as the optimal number of dimension. In our system, we take Radial Basis Function(RBF) kernel and corresponding kernel variance($\sigma_{RBF}$) is 2.5. And soft margin variable is set to 100. These parameters are selected empirical manner. And the optimal number of dimension also is selected experimentally. In order to set optimal number of dimension, we trained original UIUC car/noncar database.[2] This database contains 500 car images and 500 noncar images whose dimension is $\mathbb{R}^{20\times 50}$. Half of the images are used for training and reminders of the images are used for testing. Table 1 and figure 5 shows the result of classification rate and corresponding Bayes error rate as varying the number of reduced dimension, respectively. In figure 5, we can recognize that MMI in general is more effective than PCA. One thing noticeable is that PCA shows better Bayes error rate in 64 dimension because the amount of mutual information corresponding each basis is reduced gradually as shown in figure 6. Thus we can know that basis vector that represented by small amount of mutual information is of no use for classification task.

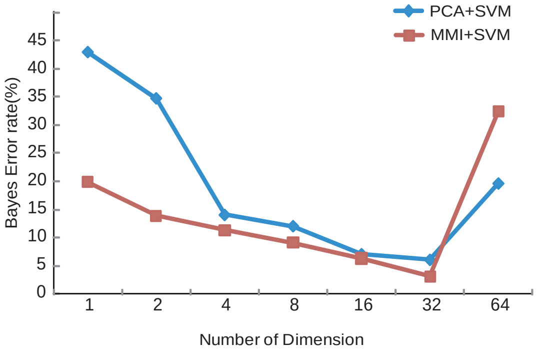

**Figure 5: Bayes error rate**

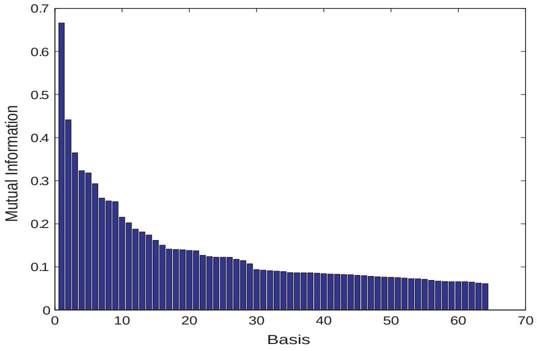

**Figure 6: The amount of mutual information corresponding basis**

## 5. SEARCHING OBJECT IN A IMAGE USING TRAINED CLASSIFIER

In this section, we discuss on an efficient detection scheme. Detection process depicted in figure 7 is referred to as an aggregation of all classification results which can be computed by moving detection window with variously scaled size. This process substantially need to so many matching trials(classification). Although we train a classifier with a fancy algorithm, detection result may be pool in real application since in order to detect a certain object in a image we may try to evaluate up to thousands of time depending on several factors such as image size, scaling factor, searching interval and so on. In short, even thought we achieve 99% of classification rate, in detection process we statistically fail to detect objects 10 times in case we perform a thousand of trials(SVM evaluation) in a image. Consequently, at the detection perspective, 99% of classification rate is not outstanding at all. Detection process thus is critical as well as training process. We hence introduce an efficient detection scheme named as Scale based Classifier Activation Map(SCAM). In this scheme, we in advance know the fact that although searching window is not exactly matched to the object, evaluated value by the trained classifier is amount of positive value. It plays a key role in sustaining our SCAM method. The SCAM method can be formulated as follow

$$M(i,j) = sig\left[\frac{1}{C}\sum_{k}^{T}\sum_{i}^{(s_k H-h)}\sum_{j}^{(s_k W-w)} y(\mathbf{n}_{ij})\right], \qquad (8)$$

where $\mathbf{n}_{ij}$ is a window vector positioned at $i, j$ in a scaled image. $T$ is a number of classification stage corresponding image scale. $s_k$ is scale parameter corresponding stage $k$. $H$ and $h$ are height of original image and the window, respectively. $W$ and $w$ are width of original image and the window, respectively. $C$ is normalization constant makes

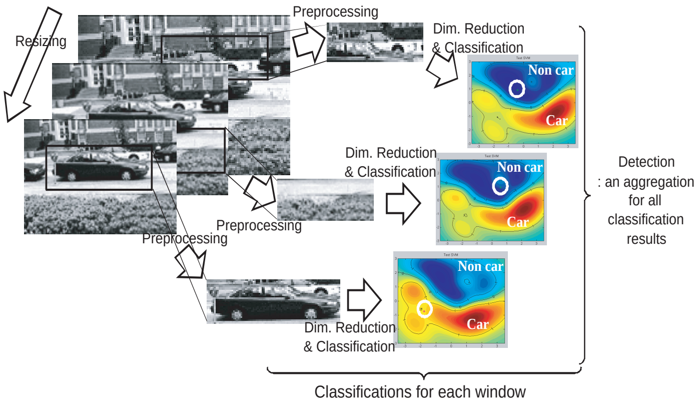

**Figure 7: An illustration of the general detection scheme**

the peak value of the cumulated map is to be one such that

$$C = \max\left[\sum_{k}^{T}\sum_{i}^{(s_kH-h)}\sum_{j}^{(s_kW-w)} y(\mathbf{n}_{ij})\right] \qquad (9)$$

And $y(\mathbf{n}_{ij})$ is a degree evaluated from trained Support Vector Machine(SVM) with respect to the window, $\mathbf{n}_{ij}$. Figure 8 illustrates an example of our detection scheme. The window, $\mathbf{n}_{ij}$ moves its position toward horizontal(indexed by $j$) and vertical(indexed by $i$) direction in order to find object in whole image with scale factor, $s_k$ depicted in second row of figure 8. The reason we need to scale the original image is that there is no way to estimate the size of the object we want to detect. Then the evaluated value by trained SVM is cumulated at the corresponding position and size of window providing that the evaluated value is positive. The third row-right of figure 8 shows an example of the result. Finally, by normalizing and thresholding the cumulated map sequentially, we can get SCAM result as depicted in third row-left of figure 8.

## 6. EXPERIMENT

We applied our detection scheme into UIUC car/non-car database[2]. This database contains 500 car and non-car images, respectively. And each training example shows 4000 dimensions. We resized all training examples to 1000 dimension that aims to reduce noisy pixel as well as dimensionality. These examples are used for training our car detection system. In test images, the UIUC car/non-car database contains 170 images with 200 cars. The detection rate metric is $\frac{\text{number of correct positive}}{\text{total number of car in the data set}}$. By using the metric, our detection scheme shows 92.5% of correct detection rate. Corresponding false positive rate is 0.05% which is calculated by the metric, $\frac{\text{number of false positive}}{\text{total number of negatives in the data set}}$. Some examples of detection result are shown in figure 10. And our method was successfully applied to automated car detection system. There are several video clips of our detection result in real application which is posted in $http://home.postech.ac.kr/\sim taey16/publication/carDetection.htm$. We demonstrated that our method practically can be used for detecting car in videos as well as still images. Several still images of detection result of automated car detection system is shown in figure 11.

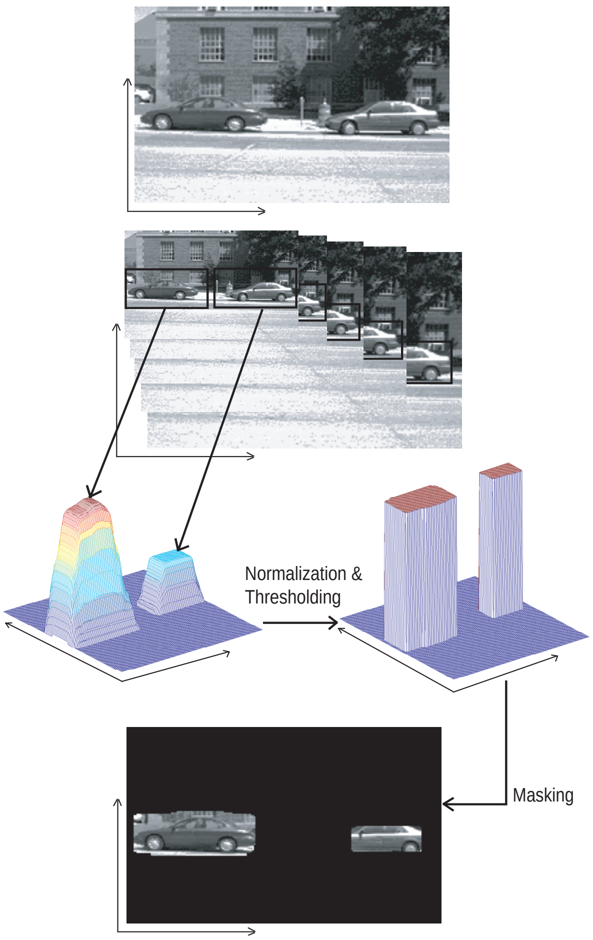

**Figure 8: An example of detection result. 1st row: Original image. 2nd row: Searching an object with variously scaled image. 3rd row-left: Cumulated searching result. 3rd row-right: SCAM. Bottom: Detection result.**

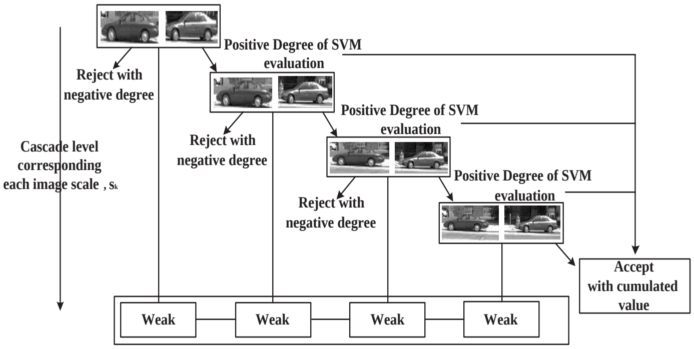

**Figure 9: Our SCAM method at the point of ensanble classifier**

## 7. DISCUSSION AND CONCLUSION

We introduced an effective searching scheme in a image. Theoretically, our algorithm is motivated by an ensanble classifier. In other words, classifier in a certain scaled image is referred to as a weak classifier conceptually. The result of weak classifiers in each scale is aggregated so that we can get more accurate result which is illustrated in figure 9. In order to achieve more desirable result, it may be good to combine both our method and component based approach introduced by [1, 9]. This component based approach basically assume that each component shows little variation comparing with whole object itself as varying camera pose or illuminance. Therefore this approach can be applied to our system successfully. In this work, we only focused on the application to the car detection however it may be applicable to detect human, since it can discriminate any objects showing separable patterns. Challenging goal of our research in feature is to detect car and human in an application simultaneously. Another thing we want to discuss is time complexity. The dimension reduction method we used shows $O(dn^2)$ time complexity. It can be a problem in real application in case that we have lots of training examples we should train. In order to reduce time complexity, the method which is used for large scale Support Vector Machine(SVM) can be computationally practicable.

## 8. ACKNOWLEDGMENTS

This work was financially supported by the Intelligent Robotics Development Program, one of the 21st Century Frontier Research & Development Programs funded by the Ministry of Commerce, Industry and Energy (MOCIE).

## 9. REFERENCES

[1] S. Agarwal and A. Awan. Learning to detect objects in images via a sparse, part-based representation. *IEEE Trans. on Pattern Analysis and Machine Intellgence*, 26(11):1475–1490, 2004.

[2] S. Agarwal and D. Roth. Car detection. *http://l2r.cs.uiuc.edu/~cogcomp/Data/Car/*, 2002.

[3] E. Bingham and H. Mannila. Random projection in dimensionality reduction: applications to image and text data. In *proceedings of the 7th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 245–250, 2001.

[4] S. Canu. Svm and kernel methods matlab toolbox. Perception Systemes et Information, INSA de Rouen, Rouen, France, 2005.

[5] R. Cendrillon and B. C. Lovell. Real-time face recognition using eigenfaces.

[6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *international Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 886–893, June 2005.

[7] Fano. Transmission of information: A statistical theory of communications. *American Journal of Physics*, 29:793–794, November 1961.

[8] Hellman and R. J. M. Probability of error, equivocation, and chernoff bound. *IEEE Trans. on Information Theroy*, 16:368–372, July 1970.

[9] A. Mohan, C. Papageorgiou, and T. Poggio. Example-based object detection in images by components. *IEEE Trans. on Pattern Analysis and Machine Intellgence*, 23(4):349–361, 2001.

[10] G. B. N. Goel and A. Nefian. Face recognition experiments with random projection. In *SPIE Defense and Security Symposium (Biometric Technology for Human Identification), Orlando, FL*, 2005.

[11] E. Osuna, R. Freund, and F. Girosi. Training support vector machines:an application to face detection. *International Conf. on Computer Vision and Pattern Recognition*, 1997.

[12] G. Qiu and J. Fang. Car/non-car classification in an informative sample subspace. In *the 18th International Conf. on Pattern Recognition*, 2006.

[13] H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(1):23–38, January 1998.

[14] O. Sidla, Y. Lypetskyy, N. Brandle, and S. Seer. Pedestrian detection and tracking for counting applications in crowded situations. In *proceedings of the IEEE International Conf. on Video and Signal Based Surveillance*, page 70, 2006.

[15] K. Torkkola. Feature extraction by non parametric mutual information maximization. *Journal of Machine Learning Research*, 3:1415–1438, 2003.

[16] UCI. Uci machine learning repository. *http://archive.ics.uci.edu/ml/datasets/Statlog+(Landsat+Satellite)*, 1993.

[17] Q. Zhu and S. Avidan. Fast human detection using a cascade of histograms of oriented gradients. In *international Conf. on Computer Vision and Pattern Recognition*, pages 1491–1498, 2006.

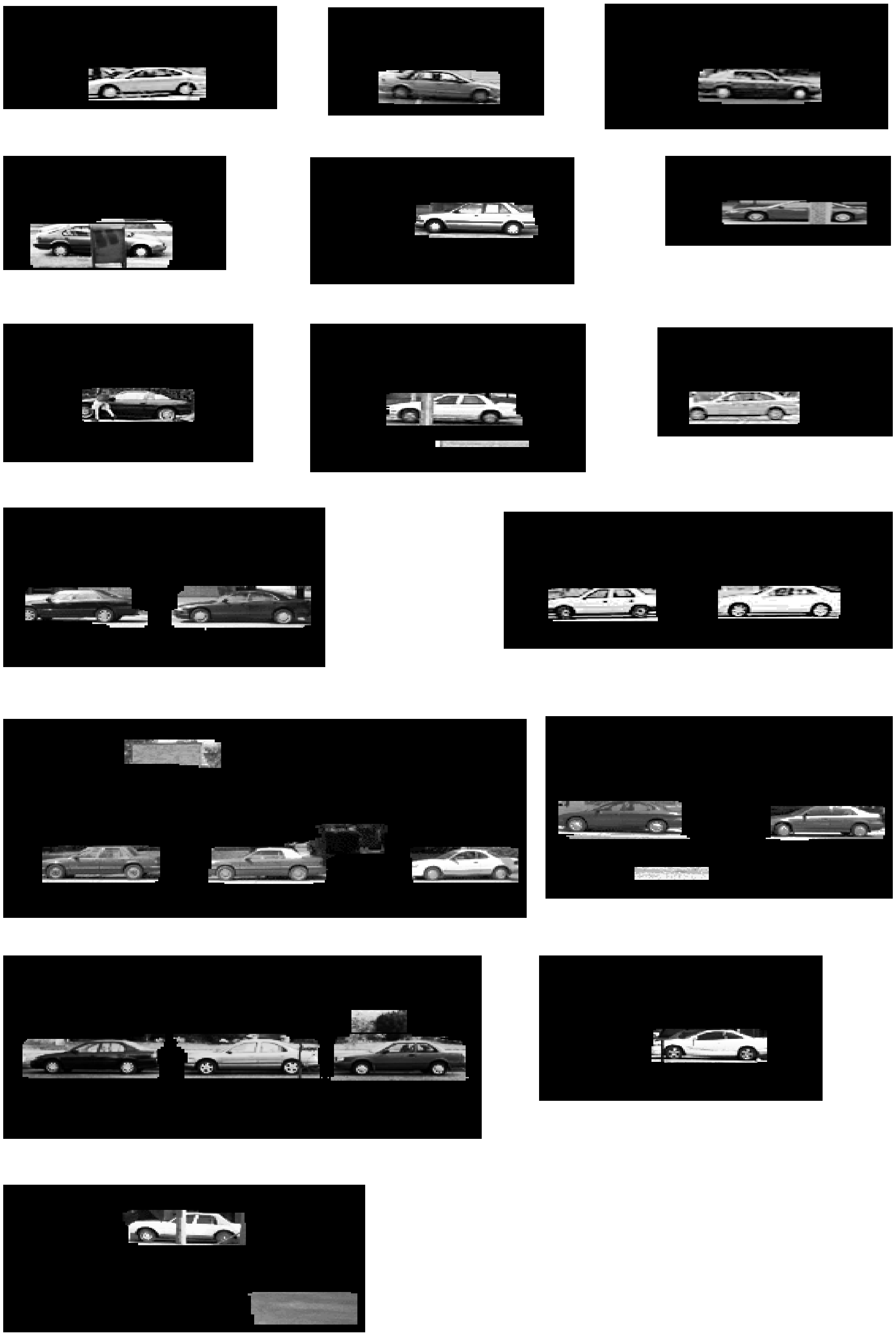

**Figure 10: UIUC image database; Examples of detection result using Scale based Classifier Activation Map(SCAM)**

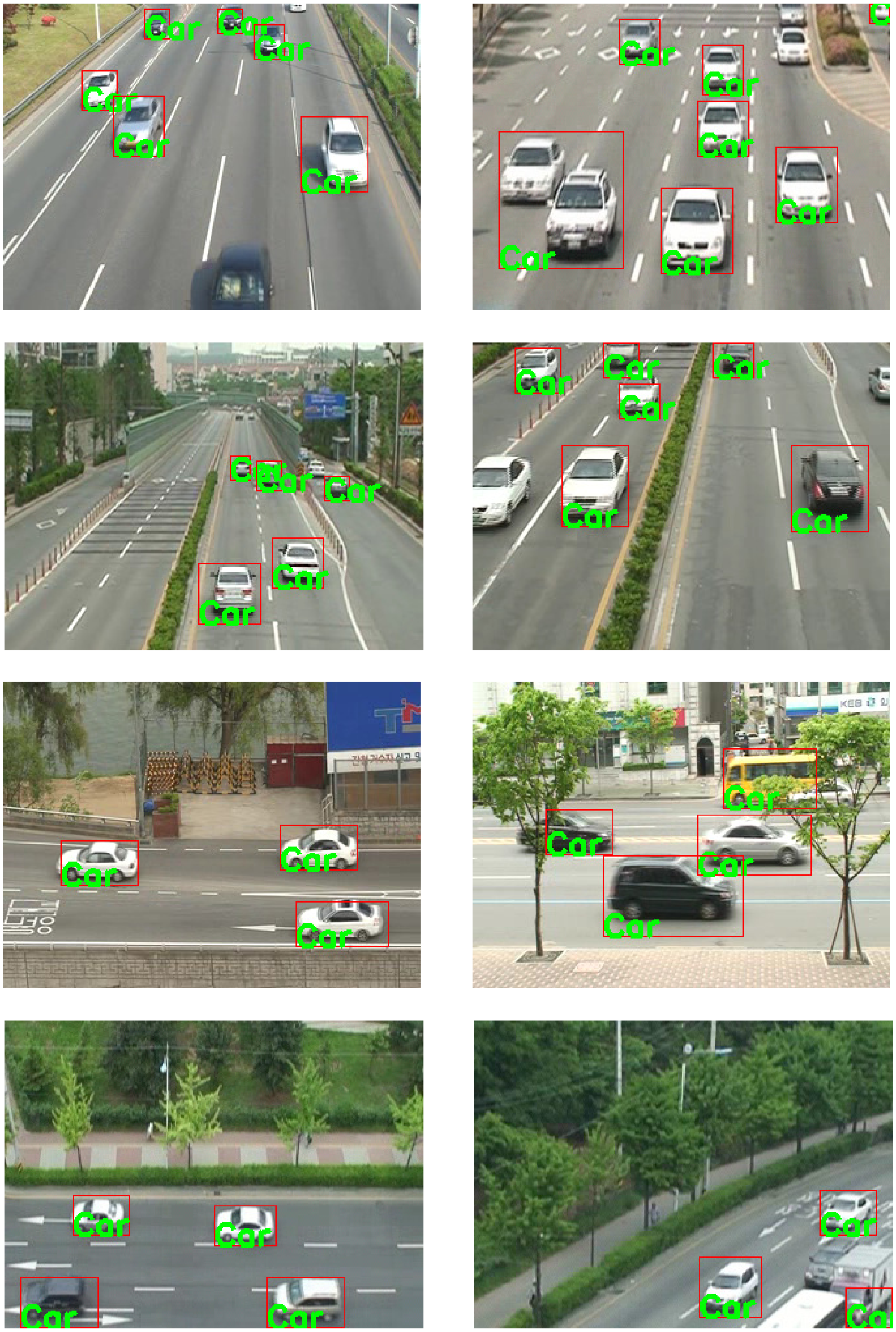

**Figure 11: Real road database; Some detection results of automated car detection system**