

Car/Non-Car Classification in an Informative Sample Subspace

Jianzhong Fang and Guoping Qiu

School of Computer Science and Information Technology, The University of Nottingham

ABSTRACT

In this paper, we present a method for data classification with application to car/non-car objects. We first developed a sample based car/non-car maximal mutual information low dimensional subspace. We then trained a support vector machine (SVM) in this subspace for the detection of cars. Using publicly available standard training and testing data sets, we demonstrated that our car detector gave very competitive performances.

1. INTRODUCTION

Car detection is a popular research topic in automatic object detection [1 - 3]. As it is common to any generic object detection, feature representation plays a key role in the success of car detection. An effective representation method should be compact and discriminative. It is desired that the representation should have low dimensionality to combat the “curse of dimensionality” problem and to improve computational efficiency. The representation should also ideally be in a space where different classes of objects are well separated.

Recently, the authors have introduced an informative subspace method for human face detection [4]. In this paper, we extend the method to other object types and show that similar idea can also be successfully applied to the detection of cars.

In this work, we use the car and non-car training and testing images from the University of Illinois at Urbana-Champaign [5]. The advantage of using this database is that both the training data and testing data are publicly available which makes it easier to compare the performances of different methods. In this database, the training set contains 550 side-view car images and 500 non-car images with a size of 100 by 40 pixels stored in PGN raw data format. The testing set contains 170 images with 200 cars in them. The scales of the cars in the testing set are approximately the same as those in the training set. The testing images are different in size and the number of cars in them. They are also stored in PGN raw data format. In the next section, we describe our car detection system,

in section 3 we present results of our method and section 4 concludes our presentation.

2. OUR CAR DETECTION SYSTEM

Our car detection system is illustrated in Fig. 1. The input image (detection window of 100 x 40 pixels) is first down-scaled by a factor of two. The purposes of down scaling are two folds. One is to reduce the dimensionality of the input and the other is to smooth the input (remove noise). We have tried a number of scaling factors and found that a down scaling factor of 2 gave the best results. Even though down scaling the detection window from 4000-d to 1000-d is a significant reduction, the dimension of the input is still too high. We therefore applied a linear transform to this 1000-d input. Even though conventional linear transforms such as PCA or LDA can be applied here, we use a maximal mutual information (MI) subspace (to be briefly explained shortly) to reduce the dimensionality of the detection window. The transformed low dimensional input vector (set to 32-d in this paper) is then fed to a support vector machine (SVM) [10] which have been trained to make a binary decision indicating whether the current input is a car or non-car object.

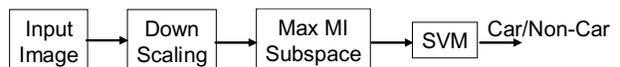


Fig. 1. A car detection system in maximal mutual information subspace

2.1 Maximal Mutual Information Subspace

In the context of object classification, Fano’s inequality [6] gives a lower bound for the probability of error (an upper bound for the probability of correct classification). The maximal mutual information subspace method [4] used Fano’s inequality in much the same way as it was used by other authors [7, 8]. The classification process can be interpreted as a Markov chain as illustrated in Fig. 2.

The probability of misclassification error in the setting of Fig. 2, $P_e = P(y \neq y')$, has the following bound [6]

$$P(y \neq y') \geq \frac{H(Y) - I(Y, F) - 1}{\log(m)} \quad (1)$$

where F is the ensemble of random variable f , and m is the number of outputs of y (number of object classes).

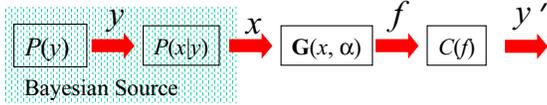


Fig. 2. Interpreting the classification process as a Markov chain, y is the object class random variable, x are the observations generated by the conditional probability density function $P(x|y)$. The observations are subjected to a transform G , which produces a new feature f from input x . The classifier C then estimates the class identity of input x as y' based on the transformed feature f .

The form of the classifier, C , has not been specified. Eq. (1) quantifies at best how well we can classify the objects using the features f . However, an upper bound of the probability of misclassification error cannot be expressed in terms of Shannon's entropy. The best one can do is to minimize the lower bound to ensure an appropriately designed classification algorithm does well. Since both m and $H(Y)$ are constants in (1), we can maximize the mutual information $I(Y, F)$ to minimize the lower bound of the probability of misclassification error. The task now becomes that of finding the transform function G that minimizes this lower bound. In this paper, we use a simple constructive approach to finding the transform bases directly from the training samples [9]. Our objective is to find a dimension reduction linear transform G that minimizes the lower bound in (1). Because the observations x and the transformed feature f and class variable y are all normally multidimensional vectors, directly estimating an optimal G that maximizes $I(Y, F)$ is computationally extremely difficult.

Assume x is an l -d column vector and f is a k -d column vector, ($k \ll l$), then $f = Gx$, G is a k (rows) by l (columns) transform matrix. Let $X = \{x_1, x_2, \dots, x_N\}$ be the N labeled training samples (car/non-car), $Y = \{y_1, y_2, \dots, y_N\}$ their corresponding class labels, $G = [g_1, g_2, \dots, g_k]^T$, and g_i be the i th transform base. Since we will find the transform bases directly from the training samples using a maximum mutual information criterion, we term our method informative sample subspace (ISS) procedure (see ISS procedure box).

To find the first transform base, we select one sample at a time, and project all other training samples onto that selected sample. The projection (a scalar) and the sample identity can be used to estimate the joint probability, which in turn can be used to estimate the mutual information of the projection and the class distribution.

The sample with projection output that maximizes the mutual information is selected as the first transform base. This base is then removed from the training sample set. All remaining samples are then made orthogonal to the first base and used as training samples to find the second transform base. The process continues until all required k bases are found. From the procedure it is not difficult to see that all k initial bases are orthonormal. Fig. 3 shows examples of 6 transform bases obtained from the training set of the car no-car data from the database of UIUC [5].

ISS Procedure Box

Informative Sample Subspace (ISS) Procedure

```

Proc. Find  $G(X, Y)$ 
for  $i = 1$  to  $k$  do
  for  $m = 1$  to  $N$  do
    for  $n = 1$  to  $N$  do
       $F(m, n) = \langle x_m, x_n \rangle$  //inner product
    End for
    Proc. Estimate joint probability  $P(Y, F(m, \bullet))$ 
    Proc. Compute  $I(m) = I(Y, F(m, \bullet))$ 
      //Compute mutual information
    End for
    If  $I(j) > I(m), \forall m$  Then  $g_i = x_j / \|x_j\|$ 
    Remove  $x_j$  and  $y_j$  from  $X$  and  $Y$  respectively
     $N = N - 1$ 
    for  $m = 1$  to  $N$  do
       $x_m = x_m - \langle x_m, g_i \rangle g_i$ 
    End for
  End for
End Proc.

```



Fig. 3 Examples of maximal mutual information subspace bases of our car detector.

3. EXPERIMENTAL RESULTS

3.1 A Simple Example

One of the reasons that mutual information (MI) has not been more widely used is because of the exponential computational complexity in the estimation of MI. Our direct sample based heuristic approach provides a computationally practicable method for using mutual information. Although our MI-based method bears some resemblance with Fisher's linear discriminant analysis (LDA) because both use labeled data, there are fundamental differences. Whilst LDA assumes unimodal equal variance Gaussian class distributions which are almost always not true for real world data, MI-based transforms do not make such a restricted assumption. Whilst LDA is not directly related to classification rates,

maximizing MI directly minimizes a lower bound of classification errors. Whilst LDA only makes use of the covariance, MI based transforms exploit higher than second order, more general statistical information. It is well known that LDA can be easily over fitted in situations with large numbers of highly correlated features and can also be under fitted when the decision boundaries are complex and nonlinear. There are reasons to believe that MI based methods could overcome such problems better than LDA because MI captures higher order, more general statistical relations between variables. Principal Component Analysis (PCA) is a widely used linear transform for dimensionality reduction. It is an optimal dimensional reduction technique in the mean square error sense. However, the transform is not necessarily suitable for pattern classification purpose.

Fig. 4 shows a 2-class problem and the first linear transform base of PCA, LDA, and the maximal mutual information subspace. It is seen that when the decision boundary is complicated, projecting the data onto LDA and PCA subspace will make the data difficult to classify. However, projecting the data onto the maximal mutual information subspace makes the data easily separable. This is conformed by building a Bayesian classifier in the 1-d subspace to classify the two classes. The classification errors in the 1-d subspaces are respectively, PCA: 22.1%, LDA: 16.2%, and the new informative subspace: 2.9%. Although these are simple toy examples, we believe they are indicative that maximal mutual information subspace may offer better performances than traditional subspaces such as PCA and LDA.

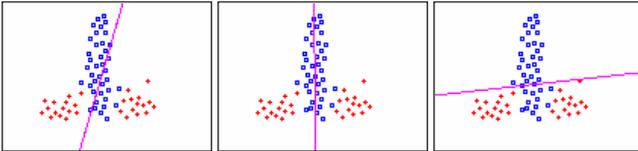


Fig. 4 Different transforms and their classification ability, the line is the first base of the respective transforms. Left: PCA projection base. Middle: LDA projection base. Right: MI projection base. Class 1 (■-blue), Class 2 (+ - red).

3.2 Car Detection (Car/Non-Car Classification)

We have applied our car detector to the standard test set of the UIUC database [5], which consists of 170 gray-scaled images with 200 cars in them. We employed exhaustive search method on the testing set. The scale of the car used in the testing was the same as the one in the training set. The 100 by 40 pixel detection window was moved 4 pixels horizontally and 2 pixels vertically each time during the evaluation. The input image patch was downscaled to 50 by 20 pixels before going through the maximal mutual information linear transform. Finally the car evaluation was taken place in the 32 dimensional maximal mutual

information subspace by the support vector machine (SVM) which was trained based on the standard method [10]. A car was counted as correctly detected if all parts of the car were enclosed in the 100 by 40 pixel window, which was a stricter criterion than that used by the data's originators [1]. The car detector evaluated 176,792 patterns over the 170 test images. In order to make comprehensive comparison we adopt three criteria as used in [1] to characterize the car detector, these criteria are Correct detection rate - Recall (R), False detection rate (F) and Precision (P):

$$R = \frac{\text{Number of correct positives}}{\text{Total number of cars in the data set}}$$

$$F = \frac{\text{Number of false positives}}{\text{Total number of negatives in the data set}}$$

$$P = \frac{\text{Number of correct positives}}{\text{Number of correct positives} + \text{Number of false positive.}}$$

The ideal detector should have 100% correct detection rate, 0% false detection rate and 100% precision. We show our car detector performance in Table 1, and show the performance of UIUC car detector in Table 2. The ROC curve of our car detector is shown in Fig. 5. From these results, we can see clearly that our car detector outperforms the original detector [1] in every quality indicator. These results demonstrate that our maximal mutual information subspace based car detector is a successful method. Fig. 6 shows example images of detection results.

Table 1: Our car detection system (Fig. 1)'s performance

Activation threshold	No. of correct detections, N	Recall (R)	No. of false detections, M	Precision (P)	False detection rate (F)
0.15	188	94.0%	45	80.6%	0.027%
0.25	184	92.0%	39	82.5%	0.023%
0.35	182	91.0%	34	84.3%	0.020%
0.45	180	90.0%	30	85.7%	0.018%
0.55	179	89.5%	25	87.7%	0.015%
0.65	174	87.0%	23	88.3%	0.014%

Table 2: Car detection performance of [1]

Activation threshold	No. of correct detections, N	Recall (R)	No. of false detections, M	Precision (P)	False detection rate (F)
0.55	181	90.5%	98	64.9%	0.09%
0.65	178	89.0%	92	65.9%	0.08%
0.75	171	85.5%	76	69.2%	0.07%
0.85	162	81.0%	48	77.1%	0.04%
0.90	154	77.0%	36	81.1%	0.03%
0.95	140	70.0%	29	82.8%	0.03%

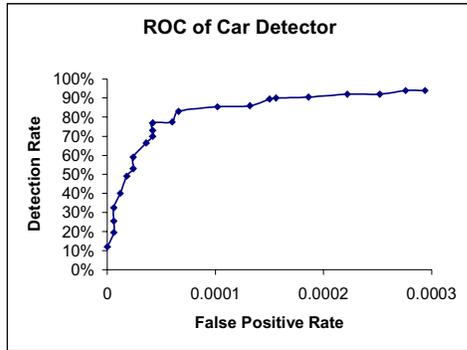


Fig. 5 ROC curve of our car detector

4. CONCLUDING REMARKS

In this paper, we have successfully developed a maximal mutual information subspace method to a popular object detection problem. This is an extension of sample based subspace method [4] and maximal mutual information subspace method [9] of our earlier works developed for human face detection. This work demonstrates that such a sample based subspace method can also be applied to more general object detection tasks.

References

1. S. Agarwal and D. Roth, "Learning a sparse representation for object detection", In Proc. ECCV2002, pages 113 -130
2. H. Schneiderman and T. Kanade, "A histogram-based method for detection of faces and cars", Proceedings ICIP 2000, International Conference on Image Processing, 2000
3. M. Vidal-Naquet, S. Ullman, "Object recognition with informative features and linear classification", ICCV 2003, Nice, France
4. J. Fang and G Qiu, "Learning an Information Theoretic Transform for Object Detection", in ICIAR'2004, Image Analysis and Recognition, A. Campilho and M. Kamel (Eds), Springer, LNCS 3211, 503-510
5. <http://l2r.cs.uiuc.edu/~cogcomp/Data/Car>
6. R. M. Fano, Transmission of Information: A Statistical Theory of Communications, MIT Press, Cambridge, MA, 1961.
7. J. W. Fisher III and J. C. Principe, "A methodology for information theoretic feature extraction", World Cogress on Computational Intelligence, March 1998
8. T. Butz, J. P. Thiran, "Multi-modal signal processing: an information theoretical framework". Tech. Rep. 02.01, Signal Processing Institute (ITS), Swiss Federal Institute of Technology (EPFL), 2002.
9. Information omitted for blind review, Proc. ICPR'2004
10. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery 2(2), pp. 121-167, 1998

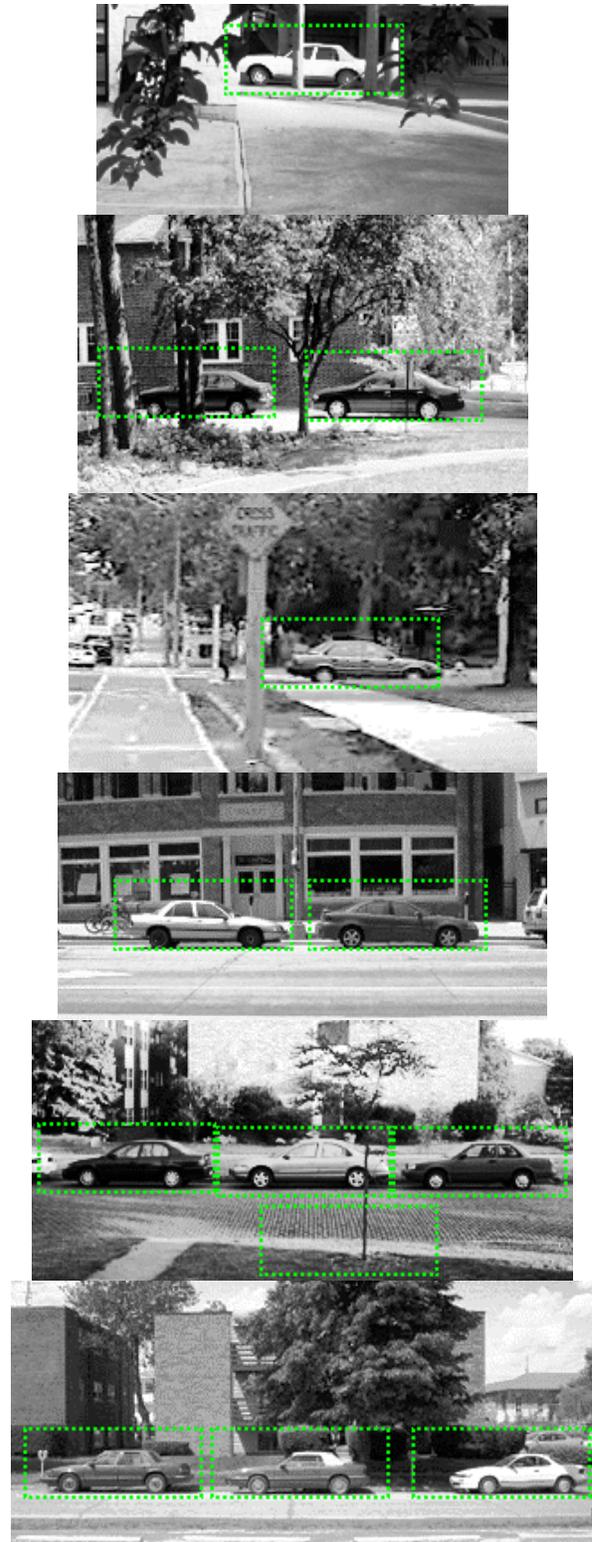


Fig. 6 Examples of car detection result of our method.