Learning Informative Features for Spatial Histogram-Based Object Detection

Hongming Zhang[1], Wen Gao[1,2], Xilin Chen[1,2], Debin Zhao[1]
[1]Department of Computer Science, Harbin Institute of Technology, Harbin,150001,China
[2]Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China
E-mail: {hmzhang, wgao, xlchen, dbzhao}@jdl.ac.cn

Abstract— **Feature extraction for object representation plays an important role in automatic object detection system. As the spatial histograms consist of marginal distribution of image over local patches, object texture and shape are simultaneously preserved by the spatial histogram representation. In this paper, we propose methods of learning informative features for spatial histogram-based object detection. We employ Fisher criterion to measure the discriminability of each spatial histogram feature and calculate features correlation using mutual information. In order to construct compact feature sets for efficient classification, we propose informative selection algorithm to select uncorrelated and discriminative spatial histogram features. The proposed approaches are tested on two different kinds of objects: car and video text. The experimental results show that the proposed approaches are efficient in object detection.**

I. Introduction

In computer vision community, object detection has been a very challenging research topic. Given an object class of interest *T* (the target) and an image *P*, object detection is the process of detecting and locating the occurrences of *T* in *P*. The main difficulty of object detection arises from high variability of appearance between objects of the same class due to different scales and poses, different lighting conditions.

Many approaches have been proposed for object detection in cluttered images. Generally speaking, these algorithms can be classified as two categories: global appearance-based approaches and component-based approaches.

Global appearance-based approaches take an object as an unit and perform classification on features generated from the entire object. Many statistical learning mechanisms are explored to identify object patterns. Neural networks are the common classification methods in detection [1,2]. Support vector machines [3,4] and Naive Bayes classifiers [5] are used to locate human faces and cars. Boosting algorithms are widely applied to detect objects, such as faces [6,7] and text [8].

Component-based methods treat an object as a collection of parts. These methods first extract some object components, and then detect objects by using geometric information. In [9], a person is represented by components such as head, arms and legs, and support vector machine classifiers are used to detect these components and decide whether a person is present. In [10], Naquest and Ullman use fragments as features and perform object recognition with informative features and linear classification. In some recent works, interest operators are first used to extract components of objects, and then perform detection by a classifier[11], or by probabilistic representation and recognition approaches[12,21].

Motivated by the observation that objects have texture distribution and shape configuration, we present spatial histogram as representation of objects. As the spatial histograms consist of marginal distribution of image over local patches, the object information about texture and shape can be encoded simultaneously. In [22], we have presented an object detection method using spatial histogram features and a hierarchical classifier that combines histogram matching and Support Vector Machine(SVM). This method is effective and efficient to detect human faces in color images.

However, this method has limitations when applying to detect other objects. (1). The spatial histogram features are predefined and quantitative analysis is missing to measure the discriminative ability of spatial histogram features. (2). As the spatial histogram features are selected as object's salient parts by hand, this method is difficult to extend to detect objects which have no influent and uniform configuration, such as side-view cars, and objects without salient parts and fixed configurations, such as text.

In order to overcome these limitations, we propose methods to automatically select informative features for spatial histogram based-object detection in this paper. First, We adopt Fisher criterion to quantitatively analyze the discriminative ability of spatial histogram features for object detection, and employ mutual information to measure features correlation. Then, we propose a training method of cascade histogram matching by automatically selecting discriminative spatial histogram features. Finally, we present a forward sequential selection algorithm to select uncorrelated and discriminative spatial histogram features for SVM classification.

The paper is organized as follows. The overview of object detection system is given in Section II. In Section III, the spatial histogram for object representation is described, and quantitative measurement of spatial histogram features is provided. Selection of informative features is presented in Section IV. Experiment results of car detection and video text detection are given in Section V. Section VI concludes this paper.

II. System Overview

We utilize an exhaustive search strategy to detect multiple object instances of different sizes at different locations in an input image. Take car detection as an example, the process

0-7803-9048-2/05/$20.00 ©2005 IEEE

of object detection in images is shown in Fig.1. The process contains three steps: image pyramid construction, object classification at different scales, and detection results fusion. In the Step 1, the original image is repeatedly reduced in size by a factor 1.2, resulting in a pyramid of images. A small window (sub window) with a certain size is used to scan the pyramid of images at different scales. It is passed to the following procedures in the Step 2. Firstly, spatial histogram features are generated from this sub window. Secondly, histogram matching and SVM classification are performed hierarchically to identify whether or not the sub window contains an object instance. In the Step 3, overlapped object instances of different scales are merged into final detection results.

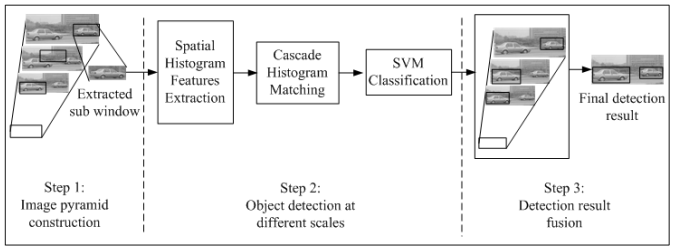

Fig. 1. Process of Object Detection in Images

III. Spatial Histogram Features

Object representation and feature extraction are essential to object detection. In this section, we describe a novel object representation combining texture and spatial structures. Specially, we model objects by their spatial histograms over local patches and extract class specific features.

A. Spatial Histograms

In our approach, a sub window contains a grey sample image with certain size. Local Binary Pattern(LBP) is used to preprocessed sample images. LBP is a relatively new and simple texture model and it has been proved to be a very powerful feature in texture classification [14]. LBP is invariant against any monotonic transformation of the gray scale. As illustrated in Fig.2, Basic LBP operator uses neighborhood values to calculate the region central pixel value.

g_1	g_2	g_3
g_4	g_0	g_5
g_6	g_7	g_8

Fig. 2. Neighborhood for LBP Computation

The 3x3 neighborhood pixels are signed by the value of center pixel:

$$s(g_0, g_i) = \begin{cases} 1 & g_i \geq g_0 \\ 0 & g_i < g_0 \end{cases}, (1 \leq i \leq 8). \quad (1)$$

The signs of the eight differences are encoded into an 8-bit number to obtained the LBP of the center pixel:

$$LBP(g_0) = \sum_{i=1}^{8} s(g_0, g_i) 2^{i-1}. \quad (2)$$

For any sample image, we compute histogram-based pattern representation as follows. First we apply variance normalization on the gray image to compensate the effect of different lighting conditions, then we use Basic Local Binary Pattern operator to transform the image into LBP image, and finally we compute histogram of the LBP image as representation. Fig.3 shows a sample image of a side-view car,its LBP images and histogram.

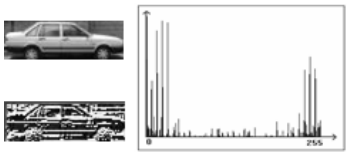

Fig. 3. An Image Sample of a Side-view Car, Its LBP Image and Histogram

It is easy to prove that histogram, a global representation of the image pattern, is invariant to translation and rotation, however, histogram is not sufficient for object detection since it does not encode spatial distribution of the object.

In order to enhance discrimination ability, we introduce spatial histograms, in which we use spatial templates to encode spatial distribution of object patterns. Each template is binary rectangle mask, shown as in Fig 4. We denote each template as $rt(x, y, w, h)$, where (x, y) is the location of the top left position of the mask and (w, h) are the width and height of the mask respectively.

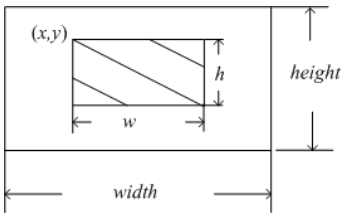

Fig. 4. Sub Window and Spatial Template

For a single spatial template $rt(x, y, w, h)$, we model sub image within the masked window by histogram. We call this kind of histogram as *spatial histogram*. For a sample image P, its spatial histogram associated with the template $rt(x, y, w, h)$ is denoted as $SH^{rt(x,y,w,h)}(P)$.

B. Object Features Extracted from Spatial Histograms

A lot of methods can be used to measure similarity between two histograms, such as quadratic distance, Chi-square distance and histogram intersection [20]. In this paper, we adopt histogram intersection for its stability and computational inexpensiveness. The similarity measurement by intersection of two histograms is defined as[13]:

$$D(H_1, H_2) = \sum_{i=1}^{k} \min(H_1^i, H_2^i), \quad (3)$$

where H_1 and H_2 are two histograms, and K is the number of bins in the histograms.

Suppose a database with n object samples and a spatial template, we represent object histogram model over the spatial template by the average spatial histogram of the object training samples, defined as:

$$SH^{rt(x,y,w,h)} = \frac{1}{n}\sum_{j=1}^{n} SH^{rt(x,y,w,h)}(P_j), \quad (4)$$

where P_j is an object training sample, and $rt(x, y, w, h)$ is the spatial template. For any sample P, we define its *spatial histogram feature* $f^{rt(x,y,w,h)}(P)$ as its distance to the average object histogram, given by

$$f^{rt(x,y,w,h)}(P) = D(SH^{rt(x,y,w,h)}(P), SH^{rt(x,y,w,h)}). \quad (5)$$

An object pattern is encoded by a spatial template set $\{rt(1), ..., rt(m)\}$, where m is the number of spatial templates. Therefore, an object sample is represented by a spatial histogram feature vector in the spatial histogram feature space:

$$F = [f^{rt(1)}, ..., f^{rt(m)}]. \quad (6)$$

As the mask can vary in positions and sizes in the scan window, the exhaustive set of spatial histogram features is very large. Therefore, the spatial histogram feature space completely encode the texture and spatial distribution of objects.

C. Discriminating Feature Analysis

Each type of spatial histogram has the discriminating ability between object and non-object pattern. To demonstrate this property, we take a spatial histogram feature of side-view car pattern as an example. The size of sample image is 100x40 pixels. The spatial template is $\{rt(40, 20, 20, 20)\}$ within the 100x40 image window locate a 20x20 mask in position (40,20). The car model over this spatial template $SH_{car}^{rt(40,20,20,20)}$ is generated by 200 car samples. The spatial histogram feature $f^{rt(40,20,20,20)}$ is the testing feature.

Fig.5 shows the testing feature's distribution over an image sample set containing 2000 car samples and 15000 non-car samples. On this feature, we use a threshold to classify car and non-car. By setting the threshold to 0.7, we retain 99.1% car detection rate with false alarm rate 45.1% and threshold 0.8 produces 93.8% detection rate with false alarm rate 12.1%.

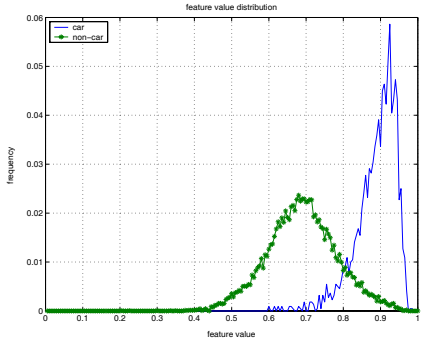

Fig. 5. Feature Distribution

We adopt Fisher criterion to measure the discriminative ability of each spatial histogram feature. For a spatial histogram feature $f_j, 1 \le j \le m$, suppose that we have a set of N samples $x_1, x_2, ..., x_N$, N_1 in the object subset labelled $\omega(1)$ and N_2 in the non-object subset labelled $\omega(2)$. The between-class scatter S_b is the distance between two classes given by

$$S_b = (m_1 - m_2)^2, \quad (7)$$

where $m_i = \frac{1}{N_i}\sum_{x\in\omega(i)} x, i \in \{1, 2\}$. The total within-class scatter S_w is defined by

$$S_w = \sum_{i=1}^{2}\sum_{x\in\omega(i)} \frac{1}{N_i}(x - m_i)^2. \quad (8)$$

Thus, the Fisher criterion of the spatial histogram feature f_j is the ratio of between-class to within-class scatter, given by

$$J(f_j) = \frac{S_b}{S_w}. \quad (9)$$

The greater Fisher criterion is, the more discriminative the spatial histogram feature is.

D. Feature Correlation Measurement

An efficient feature set requires not only each feature has strong discriminating ability, but also they are mutually independent. We employ mutual information to measure features correlation. For a spatial histogram feature, it is treated as random variable. It expresses the distance between the spatial histogram and the associated object model. For a variable X, its entropy is defined as:

$$H(X) = -\int p(x)\log_2 p(x)dx. \quad (10)$$

Given two spatial histogram features f_1, f_2, the mutual information of f_1 and f_2 is defined by

$$I(f_1|f_2) = H(f_1) + H(f_2) - H(f_1, f_2). \quad (11)$$

It is obvious $I(f_1|f_2) = I(f_2|f_1)$ and $0 \le I(f_1|f_2) \le H(f_1)$. Therefore we calculate the correlation between two features f_1 and f_2 as

$$Corr(f_1, f_2) = \frac{I(f_1, f_2)}{H(f_1)}. \quad (12)$$

Let F_s be a feature subset, we calculate the correlation between a feature $F_m \notin F_s$ and F_s as follows:

$$Corr(f_m, F_s) = \max\{Corr(f_m, f_k)|\forall f_k \in F_s\}. \quad (13)$$

IV. Learning Informative Features for Object Detection

We apply a hierarchical classification method using cascade histogram matching and SVM to object detection. Since the spatial histogram feature vector is high dimensional as mentioned in the previous section, it is crucial to get a compact and informative feature subset for efficient classification. In this section, the selection methods of informative features based on discriminability and features correlation are presented.

A. Cascade Histogram Matching

Histogram matching is a direct method for object recognition. In this method, a histogram model of an object pattern is first generated for one spatial template. If the histogram of a sample is close to the model histogram under a certain threshold, the sample is classified as an object pattern. Let P is a sample and its spatial histogram feature with one template $rt(x,y,w,h)$ is $f^{rt(x,y,w,h)}(P)$, P is classified as object pattern if $f^{rt(x,y,w,h)}(P) \geq T$, otherwise P is classified as non-object pattern. T is the threshold for classification.

Histogram matching with one spatial template is far from acceptable as an object detection system. We select most informative spatial histogram features and combine them in a cascade form to perform histogram matching. We call this classification method as cascade histogram matching. If we select n spatial histogram features $f_1, ..., f_n$ with associated classification thresholds $T_1, ..., T_n$, the decision rule of cascade histogram matching is as follows:

$$C(P) = \begin{cases} 1 & \text{object} \quad \text{if}(f_1(P) \geq T_1 \bigwedge ... \bigwedge f_n(P) \geq T_n \\ 0 & \text{non-object} \quad \text{otherwise} \end{cases}. \tag{14}$$

We measure feature's contribution by Fisher criterion and detection rate and select discriminative features to construct the cascade histogram matching.

Suppose that we have (1)spatial histogram features space $F = \{f_1, ..., f_m\}$, (2)positive and negative training sample sets: SP and SN, (3)positive and negative validation sample sets: $VP = \{(x_1, y_1), ..., (x_n, y_n)\}$ and $VN = \{(x'_1, y'_1), ..., (x'_k, y'_k)\}$, x_i and x'_i are m dimensional spatial histogram feature vectors, $y_i = 1$ and $y'_i = 0$, (4)acceptable detection rate: D. The method of training cascade histogram matching is shown in the following procedure:

(1).Initialize: $F_{select} = \phi$, $ThreSet = \phi$, $Acc(pre) = 0$;

(2).For each feature $f \in F$, compute Fisher criterion $J(f)$ on training sets SP and SN;

(3).Find the feature f' with the maximal Fisher criterion:

$$f' = \arg\max_{f_j}\{J(f_j)|f_j \in F\};$$

(4).Perform histogram matching with f' on the validation image set $V = VP \cup VN$, find a threshold θ such that the detection rate d is greater than D, i.e., $d \geq D$;

(5).Compute the classification accuracy on the set VN,

$$Acc(cur) = 1 - \frac{1}{k}\sum_{i=1}^{k}|C(x'_i) - y'_i|.$$

Here, $C(x)$ is the classification out by histogram matching with f' and θ, $C(x) \in \{0, 1\}$;

(6).If $Acc(cur) > Acc(pre) + \epsilon$ (ϵ is a small positive constant), process following steps: (6.1). $Acc(pre) = Acc(cur)$, $F_{select} = F_{select} \cup \{f'\}$, $F = F \setminus \{f'\}$, $ThreSet = ThreSet \cup \{\theta\}$, $SN = \phi$, (6.2).perform cascade histogram matching with F_{select} and $ThreSet$ on an image set containing no target objects, put false detections into SN, (6.3). goto (2);

(7).The procedure exits and returns F_{select} and $ThreSet$ for cascade histogram matching.

B. Support Vector Machine for Object Detection

Cascade histogram matching is the coarse object detection stage and it can obtain high detection rate, however, the false positive rate is still high. For the sake of improvement of detection performance, we employ Support Vector Machine(SVM) classification as the fine object detector.

A SVM [15] performs pattern recognition for a two-class problem by determining the separating hyper plane that maximum distance to the closest points of training set. In our approach, we first adopt SVM method as the evaluation classifier in the selection of informative spatial histogram features, and then use the selected feature set to train a SVM for object detection by Libsvm software [16].

By integrating discriminablity and features correlation, we use a forward sequential selection method to iteratively select a feature subset F_{select} for classification. Initially, F_{select} is set to be empty. In each iteration, this method firstly chooses an uncorrelated spatial histogram features with large Fisher criterion, then uses a classifier to evaluate the performance of the selected feature subset, and finally adds a feature which has maximum classification accuracy to F_{select}.

Suppose that we have (1)a spatial histogram features space $F = \{f_1, ..., f_m\}$, (2)a training samples set $s = \{(x_1, y_1), ..., (x_n, y_n)\}$ and a testing samples set $v = \{(x'_1, y'_1), ..., (x'_k, y'_k)\}$,where x_i and x'_i are samples with m dimensional spatial histogram feature vectors, $y_i = 0, 1$ and $y'_i = 0, 1$ for negative and positive samples respectively. The selection of feature subset F_{select} is performed as the following procedure:

(1).Find f^* with maximum Fisher criterion, $F_{select} = \{f^*\}$ and $F_{ori} = F \setminus f^*$;

(2).Classification accuracy $Acc(pre) = 0$;

(3).For each feature $f \in F_{ori}$, compute Fisher criterion $J(f)$ and $Corr(f, F_{select})$;

(4).Compute $Thre$ as follows:

$$\begin{cases} MinCorr = \min\{Corr(f, F_{select})|f \in F_{ori}\} \\ MaxCorr = \max\{Corr(f, F_{select})|f \in F_{ori}\} \\ Thre = MinCorr * (1 - \alpha) + max * \alpha \end{cases},$$

where $0 < \alpha < 1$, we choose $\alpha = 0.2$ in experiments;

(5).Find $f' \in F_{ori}$ with large Fisher criterion as below:

$$f' = \arg\max_{f_j}\{J(f_j)|Corr(f_j, F_{select}) \leq Thre\};$$

(6).Train a evaluation classifier h on the training examples set s, using f' and F_{select};

(7).Evaluate the classifier h on the testing examples set v, and compute the classification accuracy:

$$Acc(cur) = 1 - \frac{1}{k}\sum_{i=1}^{k}|h(x'_i) - y'_i|.$$

Here, $h(x)$ is the classification out by the classifier h using f' and F_{select}, $h(x) \in \{0, 1\}$;

(8).If $Acc(cur) > Acc(pre) + \epsilon$ (ϵ is a small positive constant), $Acc(pre) = Acc(cur)$, $F_{select} = F_{select} \cup \{f'\}$, $F_{ori} = F_{ori} \setminus \{f'\}$, goto (3); Otherwise, the procedure exits and returns F_{select} that contains the selected features.

V. Experimental Results

In order to evaluate the effectiveness of the proposed approaches, we conduct experiments of two different object detection tasks. One is to detect side-view car, which has semi-rigid structure with special componential configuration. The other is text detection in video frames. Text region is mainly a texture pattern without any obvious componential structure.

A. Car Detection

Side-view car consists of distinguishable parts such as wheels, car doors, and car windows. These parts are arranged in a relatively fixed spatial configuration. Unlike human faces, side-view cars have enormous changes in configurations because of various design styles.

We build a training image database with 2725 car samples and 14968 non-car samples, each 100x40 pixel in size. 500 car sample images are from the training image set from the UIUC Image Database for Car Detection [19]. Other car images are collected from video frames and websites. We also construct a validation set containing 1225 car images and 7495 non-car images for training cascade histogram matching and selection of informative classification features.

The exhaustive spatial template set within 100x40 image window is very large, 3594591. However, this spatial template set is overcomplete, and most spatial templates are with small and meaningless size or mutual overlapped. To reduce redundant and meaningless spatial templates, the mask is moved in steps of size 5 pixels in the horizontal and vertical directions and only those spatial templates, whose masks are multiple times the size of 10x10, are used in car detection. In total, 270 spatial templates in a 100x40 image sample are evaluated to extract spatial histogram features. In our experiment, 15 spatial templates(see Fig.6.) are learned for cascade histogram matching and 25 are learned for SVM classification with RBF kernel function.

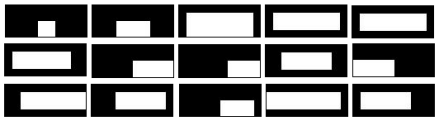

Fig. 6. Selected 15 Spatial Histogram Features for Car Detection

We test our system on a test image set from the UIUC Image Database for Car Detection [19]. The test set consists of 170 images containing 200 cars. Our system can detect side-view cars under complex backgrounds. In Fig.7, some car detection examples are given. The examples demonstrate that our approach can handle multiple cars with complex backgrounds. The ROC Curve is shown in Fig.8. The test results on the image set are shown in Table I. Compared with the system of [11,12,21], our approach achieves comparable performance with high detection rate and low false alarms.

B. Video Text Detection

Text detection is the process of detecting and locating regions that contain texts from a given image. We apply the proposed approach to detect text in video frames. We define a text block pattern as an image window 50x20 in size and construct a text region classifier using spatial histogram features. We detect text regions by two steps. First, we scan the image at multiple scales by the text region classifier to product a text region map. Second, we segment text regions into distinct text lines using vertical segmentation algorithm similar to [17].

We build a training database with 1936 text images extracted form video frames and 12313 non-text images, each 50x20 pixels in size. Similar to car detection experiments, 130 spatial templates are evaluated to extract spatial histogram features. 23 spatial templates are learned for cascade histogram matching and 32 are learned for a RBF kernel SVM.

Our system was tested on a video text set from [18]. These images are extracted from the MEPG-7 Video Content Set.

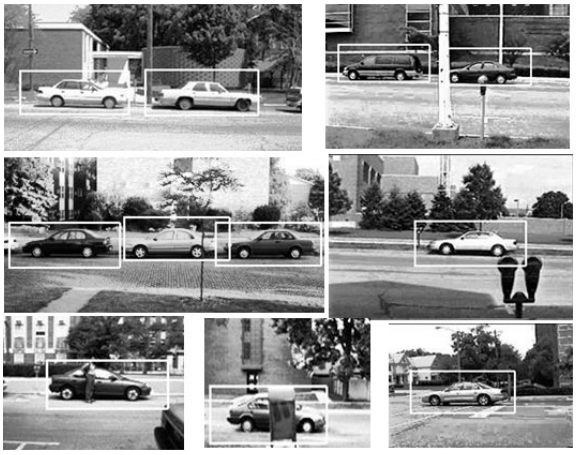

Fig. 7. Examples of Car Detection (UIUC Test Set)

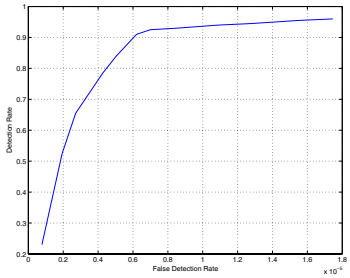

Fig. 8. The ROC Curve Obtained on UIUC Car Detection Test Set

TABLE I
Testing Results of Car Detection

	[11]	Ours	[12]	[21]
No. of correct detections,*TP*	183	193	—-	—-
No. of false detections,*FP*	557	45	—-	—-
Detection rate,*TP/200*	91.50%	96.5%	—-	—-
Precision,*TP/(TP+FP)*	24.73%	81.10%	—-	—-
Equal Error Rate	77.0%	90.5%	88.5%	91.0%

There are in total 128 human-recognizable textboxes in 45 frames. On the test image set, the system correctly detected 116 text lines and produced 16 false alarms. The testing result is listed in Table II. The correct detection rate is 90.63% and the precision is 87.88%. These results prove that the proposed object detection approach is effective in detecting video text.

TABLE II
TESTING RESULTS OF TEXT DETECTION

No. of text lines,T	128
No. of correct detections,TP	116
No. of false detections,FP	16
Detection rate,TP/T	90.63%
Precision,$TP/(TP+FP)$	87.88%

The final text detection system can detect text regions under complex background in video frames. In Fig.9, some video text detection examples are given. The examples demonstrate that our approach can handle multiple lines of text regions with different sizes.

Fig. 9. Text Detection Examples on the Video Text Set

VI. CONCLUSIONS

In this paper, we present learning approaches to select informative features for spatial histogram-based object detection. Extensive experiments have been carried out on two different kinds of detection tasks: car detection and video text detection.

In summary, this paper contains three main contributions. (1)We provide quantitative analysis of spatial histogram features. Fisher criterion and mutual information are employed to measure feature discriminative ability and features correlation. (2)The second contribution is a method of training cascade histogram matching and a method of selection informative feature subset for efficient SVM object classification. (3)We extend our approach from [22] to detect generic objects. The object detection framework are applied to objects with parts, such as cars, and those without a fixed part-based configuration, such as text in video frames. The experimental results show that learned spatial histogram features are discriminative representation for object detection and the proposed approach is efficient and robust for object detection.

ACKNOWLEDGMENT

This research is partially supported by National Nature Science Foundation of China (No. 60332010), "100 Talents Program" of Chinese Academy of Sciences, ShangHai Municipal Sciences and Technology Committee (No. 03DZ15013), and ISVISION Technologies Co., Ltd. The authors would like to thank Mr. Qixiang Ye for his help on video text detection.

REFERENCES

[1] H.A. Rowley, S. Baluja, and T. Kanade, "Neural Network-Based Face Detection," *IEEE Translations on Pattern Analysis And Machine Intelligence*, 1988,vol. 20(1), pp. 29–38.
[2] C. Garcia, M. Delakis,"Convolutional Face Finder: A Neural Architecture for Fast and Robust Face Detection," *IEEE Translations on Pattern Analysis And Machine Intelligence*, 2004,vol. 26(11),pp. 1408–1423.
[3] E. Osuna, R. Freund, and F. Girosi,"Training Support Vector Machines: an Application to Face Detection," *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 1997,pp. 130–136.
[4] C.P.papageprgiou,T.Poggio,"A training object system: car detection in static images," MIT AI Memo No.180, 1999,Octorber.
[5] H. Schneiderman, T. Kanade. "A Statistical Method for 3D Object Detection Applied to Faces and Cars,"*IEEE Conference on Computer Vision and Pattern Recognition*, 2000.
[6] P. Viola, M. Jones. "Robust Real Time Object Detection," *IEEE ICCV Workshop on Statistical and Computational Theories of Vision*, 2001.
[7] S.Z. Li, et al. "Statistical Learning of Multi-View Face Detection," *Proc. of the 7th European Conf. on Computer Vision*, 2002.
[8] X.R.Chen, A.Yuille. "Detecting and Reading Text in Natural Scenes," *Proc. of the IEEE International Conference. On Computer Vision and Pattern Recognition*, 2004.
[9] A.Mohan, C. Papageorgiou, T.Poggio. "Example-Based object detection in images by components," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001,23(4),pp. 349–361.
[10] M.V.Naquest,S.Ullman. "Object Recognition with Informative Features and Linear Classification," *Proceedings of 9th International Conference on Computer Vision*, 2003,pp. 281–288.
[11] S.Agarwal, A.Awan, D.Roth. "Learning to Detect Objects in Images via a Sparse, Part-Based Representation,"*IEEE Translations on Pattern Analysis And Machine Intelligence*, 2004,vol. 26(11), pp. 1475–1490.
[12] R.Fergus,P.Perona,A.Zisserman."Object Class Recognition by Unsurpervise Scale-Invariant Learning,"*Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2003,vol.(2)pp. 264–271.
[13] M. Swain, D. Ballard. "Color indexing," Int.J.Computer Vision, 1991,vol.(7),pp. 11–32.
[14] T. Ojala, M. Pietikäinen, T. Mäenpää. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Macine Intelligence*, 2002,vol. 24(7),pp. 971–987.
[15] V.Vapnik. *Statistical Learning Theory*. Wiley, New York, 1998.
[16] C.C. Chang, C.J.Lin. *Libsvm-a library for Support Vector Machines*. www.csie.ntu.edu.tw/ cjlin/libsvm,2004.
[17] R. Lienhart, A. Wernicked,."Localizing and segmenting text in images and videos, "*IEEE Transactions on Circuits and Systems for Video Technology*, 2002,vol. 12(4), pp. 236–268.
[18] X.S.Hua, W.Y.Liu, H.J.Zhang. "An Automatic Performance Evaluation Protocol for Video Text Detection Algorithms," *IEEE Transactions on Circuits and Systems for Video Technology*, 2004,vol. 14(4), pp. 498–507.
[19] UIUC Image Database for Car Detection. http://l2r.cs.uiuc.edu/ cogcomp/Data/Car/, 2004.
[20] Bernt Schiele."Object Recognition using Multidimensional Receptive Field Histograms,"PhD Thesis, I.N.P.Grenoble. English translation.1997.
[21] B.Leibe,B.Schiele."Scale-Invariant Object Categorization using a Scale-Adaptive Mean-Shift Search,"*Proceedings of DAGM'04 Annual Pattern Recognition Symposium, Springer LNCS, 3175*,2004,pp. 145–153.
[22] H.M.Zhang,D.B.Zhao."Spatial Histogram Features for Face Detection in Color Images,"*5th Pacific Rim Conference on Multimedia, Lecture Notes in Computer Science 3331*,2004,pp377-384.