ADAPTIVE FRÉCHET KERNEL BASED SUPPORT VECTOR MACHINE FOR TEXT DETECTION

Shiyan Hu, Minya Chen

Department of Computer and Information Science
Polytechnic University
Brooklyn, NY 11201
shu@cis.poly.edu, mchen@vision.poly.edu

ABSTRACT

In this paper, a novel general paradigm for text detection using support vector machine (SVM) is proposed. Unlike prevailing techniques in the literature, our adaptive SVM incorporates information from each input image. In addition, for better classification results and higher efficiency, a novel kernel called the Fréchet Kernel is presented for SVM classification. The adaptive SVM aims to serve as a general paradigm to improve the prevailing techniques. In the experiment, we apply the paradigm to a simple algorithm and successfully obtain a new competitive method for text detection.

1. INTRODUCTION

Recent developments in acquisition and storage technology facilitate large collections of digital images and videos. It remains an active research area to effectively and efficiently index and retrieve information from multimedia database. Due to the characteristics of text in image and video, such as low resolution and high complexity of background, it is still a challenging problem to accurately detect regions that contain text and yet keep the false alarm to a minimum.

To date, techniques in video text detection can be broadly categorized into three major groups [1]: learning-based, connected component-based, and texture-based methods. Learning based approaches adopt modern machine learning techniques, such as support vector machine (SVM) and neural network (NN) learning, for text and non-text block classification. Connected component-based approaches apply connected component analysis to the image and non-text is then filtered by geometric properties. Texture-based approaches view video text as regions that are composed of special texture patterns. Some low level image features such as edge gradients and corners are utilized to model and detect textual patterns. Despite numerous techniques in the literature, there is little work on designing methods which can serve as general paradigms to improve other techniques.

In this paper, such a paradigm is presented and it uses the support vector machine (SVM) technique. Unlike the prevailing methods in the literature, our SVM incorporates information from the input image. In addition, for better classification results as well as higher efficiency, a novel kernel called the Fréchet Kernel is presented for SVM classification. Our general paradigm of image specific SVM aims to improve numerous prevailing techniques in significantly increasing the detection rate while introducing only slightly more false alarm. In the experiment, we apply the paradigm to a simple toy algorithm and obtain a new competitive method for text detection.

2. ADAPTIVE SVM PARADIGM

Applying SVM for text detection is not a new idea - there are successful studies in the literature such as [2, 3, 4]. Loosely speaking, all of such methods use the SVM in the way that an SVM is first trained through using some blocks collected from a general set of images in two (i.e., text and non-text) categories, then the blocks from input images are classified by the trained SVM. Due to the excellence of SVM in classifying similar but different data, this type of approaches usually lead to high detection rate. The main drawback is that the false detection rate is usually high if no postprocessing techniques are further applied, refer to [2]. Clearly, to depress the false alarm while keeping the very high detection ratio is critical.

Unlike our SVM-based alternatives, we try to incorporate the image specific information into an SVM so as to make the SVM especially good at classifying blocks of the very image. To this effect, we could first apply a generic SVM trained by general image data, then select some resulting text and non-text blocks, and some general training data, to train a new SVM. With this adaptive SVM, we can generally achieve higher detection rate, however, false rate may be even higher. Therefore, it is more useful to apply our paradigm to improve the detection rate of an algorithm with

0-7803-8874-7/05/$20.00 ©2005 IEEE

low false rate. We explore this idea and obtain some inspiring results, refer to Section 4. The details of our paradigm are elaborated as follows.

Note that an adaptive SVM corresponds to exactly one image. We first place a grid on an image processed by any prevailing method with low false rate and partition it into small fixed-size blocks (of size 10×10 in our experiment). The training blocks for the image are then selected such that p percent of the featured blocks come directly from the partitioned image, while others are from the general training data. In addition, the multi-resolution strategy is explored, that is, training blocks are first normalized to a fixed size, then classification is carried out at different scales which enables detecting a variety of font sizes. Refer to Section 4 for details. For each image, we are ready to train the SVM and then classify the blocks of that image. In this way, the classification results can be improved since the new customized SVM incorporates some image specific information as well as general image information.

The SVM exploited in this paper is not a conventional one. Although SVMs have become a very successful discriminative approaches to pattern classification, the straightforward application of it to large sequence of vectors is often ineffective or inefficient [5]. As is well known, much of classification power of SVM lies in the choice of kernels and actually the study of kernels has attracted lots of research attention recently. Standard kernels such as linear and Gaussian kernels do not make full use of information from image data [5]. Therefore, a lot of new kernels have been proposed in the literature such as [6, 5, 7]. Some of the kernels have been successfully applied to image processing and shown to be superior to standard kernels [5]. In this paper, we propose a novel kernel, the *Fréchet Distance* based Kernel (or Fréchet Kernel in short), which can be viewed as an improved version of the *Kullback-Leibler divergence* based kernel originally proposed in [5], for SVM classification. The purpose for proposing the Fréchet kernel is mainly for high efficiency, since we need to train an SVM for each image, which is time consuming.

3. FRÉCHET DISTANCE BASED KERNEL

The new kernel proceeds in the same way as [5]: we start with estimating the parameters θ_i of a PDF for each image $X = \{x_1, x_2, \ldots, x_m\}$. After $p(x|\theta)$ is estimated, the kernel computation is reduced from the original sequence space to the PDF space: $K(X_i, X_j) \rightarrow K(p(x|\theta_i), p(x|\theta_j))$. We are to map the input space X_i to a new feature space θ_i. [5] uses the symmetric Kullback-Leibler divergence to define the kernel distance in the new feature space, however, this type of distance is not scalable and computationally expensive (see [5]). We adopt a more scalable and efficient metric in this paper, namely, the Fréchet distance. Before computing the Fréchet distance, we need to first discretize the distribution curves. Evidently, the distance is scalable by discretizing a curve to different fine scales. Now it remains to present an efficient algorithm for computing the Fréchet distance.

The study of computing standard Fréchet distance, which is considered to be better than the well-known *Hausdroff distance*, can be found in [8]. They are able to compute the Fréchet distance between two polygonal curves in time $O(pq \log pq)$, where p and q are the number of segments on the polygonal curves. However, since they use the parametric search technique, the algorithm is too involved and not practical. Therefore, in this paper, we consider the *"coupling Fréchet distance"* [8] which is a good approximation to the standard Fréchet distance and can be computed in $O(pq)$ time by a very simple algorithm in [8]. For completeness, we include the details in [8] as follows.

Following [8], define a curve as a continuous mapping $f : [a, b] \rightarrow V$, where $a, b \in \mathcal{R}$ and $a \leq b$ and (V, d) forms a metric space. Given $p(x|\theta_i) : [a, b] \rightarrow V$ and $p(x|\theta_j) : [a', b'] \rightarrow V$, the Fréchet distance between them is defined in [8] as

$$D_F(p(x|\theta_i), p(x|\theta_j)) = \quad \inf_{t\in[0,1]} \max d(p(x|\theta_i)(\alpha(t)), p(x|\theta_j)(\beta(t)))$$

where α and β are two arbitrary continuous nondecreasing functions from $[0, 1]$ onto $[a, b]$ and $[a', b']$, respectively. For practical purpose, we need to approximate a curve by a polygonal one to approximate the Fréchet distance between two arbitrary curves. A "polygonal curve" [8] is defined as a curve $P : [0, n] \rightarrow V$, where n is a positive integer, such that for each $i \in \{0, \ldots, n-1\}$, the restriction of P to the interval $[i, i+1]$ is affine. We denote the sequence $(P(0), \ldots, P(n))$ of endpoints of the line segment of P by $\sigma(P)$. Let P and Q be polygonal curves and $\sigma(P) = (u_1, \ldots, u_p)$ and $\sigma(Q) = (v_1, \ldots, v_q)$ be the corresponding sequences. A "coupling" [8] L between P and Q is a sequence $(u_{a_1}, v_{b_1}), \ldots, (u_{a_m}, v_{b_m})$ of distinct pairs from $\sigma(P) \times \sigma(Q)$ such that $a_1 = 1, b_1 = 1, a_m = p, b_m = q$, and for all $i = 1, \ldots, q$, we have $a_{i+1} = a_i$ or $a_{i+1} = a_i + 1$, and $b_{i+1} = b_i$ or $b_{i+1} = b_i + 1$. We denote by $|L|$ the length of the longest link in L, i.e., $|L| = \max_{i=1,\ldots,m} d(u_{a_i}, v_{b_i})$. Given polygonal curves P and Q, their coupling Fréchet distance is defined as [8]

$$D_{cF} = \min\{|L| \mid L \text{ is a coupling between P and Q}\}$$

Since a coupling with maximal edge r gives a way of walking around P and Q with leash at most r, one easily sees $D_F(P, Q) \leq D_{cF}(P, Q)$, that is, the coupling Fréchet distance is an upper bound for the standard Fréchet distance (see [8]). The main advantage of the coupling measure is that we do not need to carry out parametric search paradigm and the measure can be efficiently computed by a simple algorithm proposed in [8] as follows. We take two polygonal curves $P = (u_1, \dots, u_p)$ and $Q = (v_1, \dots, v_q)$ as input and compute the coupling Fréchet measure between them. In the algorithm, we first initialize (all elements of) a $p \times q$ matrix CA to -1, then call the following recursive function with parameters p and q.

```
Function recursive coupling $(i, j)$ [8]
  if CA$(i, j) >$ -1 then return CA$(i, j)$
  CA$(i, j)$= $\infty$
  if $i = 1$ and $j = 1$ then CA$(i, j)$= $d(u_1, v_1)$
  if $i > 1$ and $j = 1$ then CA$(i, j)$= $\max\{c(i-1, 1), d(u_i, v_1)\}$
  if $i = 1$ and $j > 1$ then
    CA$(i, j)$= $\max\{c(1, j-1), d(u_1, v_j)\}$
  if $i > 1$ and $j > 1$ then CA$(i, j)$= $\max\{\min(c(i-1, j),$
    $c(i-1, j-1), c(i, j-1), d(u_i, v_j))\}$
return CA$(i, j)$
```

A straightforward analysis of the above algorithm gives the running time $O(pq)$.

4. EXPERIMENTAL RESULTS

We first apply the general SVM to the image set. We then train the image specific SVM by information from the resulting image as discussed in Section 2. p is set to 0.5 in our experiment. We have one SVM per image and each image is then classified by the customized SVM. However, we found out that the detection rate is raised but the false detection is also raised (Refer to Table 1, adaptive SVM after SVM). Therefore, simply applying the paradigm to a method with high false alarm (e.g., the general SVM method in our case) might not produce good results, provided no postprocessing techniques such as the one used in [2] is applied.

We now apply the general paradigm to a toy text detection algorithm which is a simplified version of the detection algorithm investigated in [9] and has a median detection rate and a low false detection rate. Applying our general paradigm largely increases the detection rate while only slightly increases the false detection rate. Refer to Figure 1, where the results by the general SVM are also shown for comparison. The detection rate and false rate for the toy algorithm, general SVM and adaptive SVM are summarized in Table 1. Clearly, by just applying our general paradigm to the simple toy, we have already obtained a competitive method for text detection. Furthermore, the adaptive SVM is capable of detecting text under a variety of environments, such as complex backgrounds as well as simple background but with text of different fonts.

For completeness, we include here a brief description of the toy program, which has a low false rate. The toy first computes for each pixel a measure called MGD (*Maximum Gradient Difference*) [9], which is the maximum difference of horizontal gradients in a 1×20 window. Text regions usually have higher MGD values than non-text regions. The MGD values is then thresholded with the value 130, which is experimentally determined to depress false alarms. On the thresholded high MGD pixels, connected component analysis is applied and those regions that have inappropriate geometrical properties are removed. Note that our purpose is to decrease as much false detection rate as possible, so the criteria for the filtering process is quite strict yet simple. The further details of the toy are omitted due to space limitation. Refer to Table 1 for its detection rate as well as the false rate. Note that we also compare the new method to the one proposed in [9] where MGD is fully explored.

Table 1. Comparison of detection rate and false rate by the toy, general SVM and adaptive SVM.

	Detection rate	False rate
Toy	67.5%	1.2%
General SVM	96.5%	85.7%
Adaptive SVM after SVM	99.1%	113.5%
Adaptive SVM after Toy	96.3%	4.9%
Method in [9]	93.6%	3.5%

Table 2. Comparison of running time by using standard RBF kernel based SVM and the Fréchet Kernel based SVM.

	Time per image (sec)
RBF kernel based SVM	28.9
Fréchet kernel based SVM	12.8

Recall that we need one SVM per image. Training so many SVMs is a time-consuming task. However, due to the proposed efficient Fréchet Kernel based SVM, we are able to complete the task within considerably shorter time. As indicated in Table 2, applying our Fréchet kernel doubles the speed of the common kernel.

Finally, it is worth reminding readers that our paradigm explores the multi-resolution strategy. Let us make a few comments on it. In the case of a big font size and a small block size, the block close to the true stroke is usually identified by the toy as a text block, which is reasonable. Then, if such blocks are chosen as text training blocks, we will have some wrong training data for SVM. The consequence is the misclassification of some non-text blocks to text blocks (refer to the red cloud in Figure 2 (a)), and thus false rate is increased. To solve this problem, we resort to the multi-resolution strategy, i.e., we first normalize the training data to a fixed size so as to ensure that every training block captures the essential property of text. In the testing phase, SVM is applied at various scales of the test image to ensure that text of various fonts can be detected. Further details are omitted here due to space limitation. Refer to Figure 2 (b)

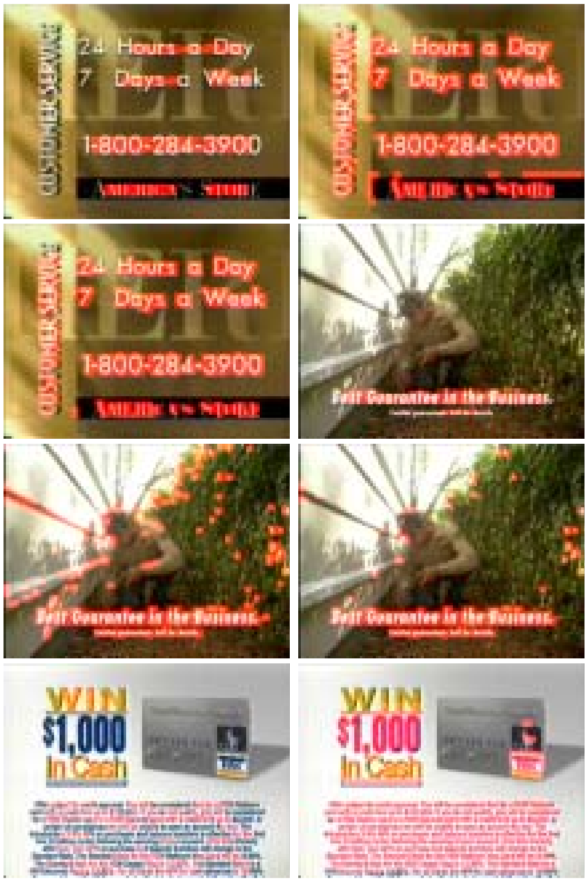

Fig. 1. Indexed from left to right and from top to bottom, the first six images are the detection results produced by the toy, general SVM and adaptive SVM after toy. The last two images are produced by general SVM and adaptive SVM after toy.

for the detection result with multi-resolution strategy.

5. CONCLUSIONS

The main contribution of this paper is two-fold. First, we propose a new SVM based paradigm for improving the prevailing text detection techniques. Second, we propose a new computationally efficient kernel, which works especially well for SVM classification in image processing. The experimental results are encouraging: applying the paradigm to a simple toy gives us a new competitive technique for text detection. The future work is to apply the paradigm to other prevailing techniques.

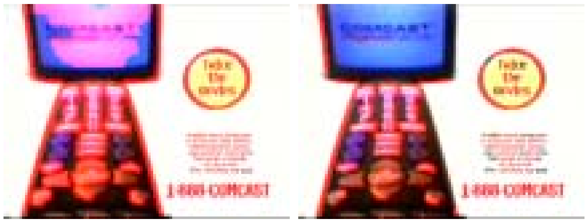

Fig. 2. (a) Detection results without multi-scale (left), and (b) with multi-scale (right).

6. REFERENCES

[1] J. Xi, X.-S. Hua, X.-R. Chen, W. Liu, and H.-J. Zhang, "A video text detection and recognition system," *IEEE International Conference on Multimedia and Expo (ICME)*, 2001.

[2] K. Kim, K. Jung, and H. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1631–1639, 2003.

[3] K. Kim, K. Jung, S. Park, and H. Kim, "Support vector machines for texture classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 11, pp. 1542–1550, 2002.

[4] ——, "Support vector machine-based text detection in digital video," *Pattern Recognition*, vol. 34, no. 2, pp. 527–529, 2001.

[5] P. Moreno, P. Ho, and N. Vasconcelos, "A kullback-leibler divergence based kernel for SVM classification in multimedia applications," *Advances in Neural Information Processing Systems (NIPS)*, 2003.

[6] T. Jaakkola, M. Diekhans, and D. Haussler, "Using the fisher kernel method to detect remote protein homologies," *Proceedings of the Seventh International Conference on Intelligent Systems for Molecular Biology*, vol. 149-158, 1999.

[7] C. Leslie and R. Kuang, "Fast kernels for inexact string matching," *Proceedings of the 16th Annual Conference on Computational Learning Theory*, pp. 114–128, 2003.

[8] H. Alt and M. Godau, "Computing the fréchet distance between two polygonal curves," *International Journal of Computational Geometry and Applications*, vol. 5, no. 75-91, 1995.

[9] E. Wong and M. Chen, "A new robust algorithm for video text extraction," *Pattern Recognition*, vol. 36, no. 6, pp. 1397–1406, 2003.