HISTOGRAM INTERSECTION KERNEL FOR IMAGE CLASSIFICATION

Annalisa Barla[1,2], *Francesca Odone*[2], *Alessandro Verri*[1,2]

(1) DISI, UniversitaÁdi Genova, Genova (I)
(2) INFM, UdR di Genova, Genova (I)

ABSTRACT

In this paper we address the problem of classifying images, by exploiting global features that describe color and illumination properties, and by using the statistical learning paradigm. The contribution of this paper is twofold. First, we show that *histogram intersection* has the required mathematical properties to be used as a kernel function for Support Vector Machines (SVMs). Second, we give two examples of how a SVM, equipped with such a kernel, can achieve very promising results on image classi£cation based on color information.

1. INTRODUCTION

High-level scene understanding is often addressed via image classi£cation. In many cases images are represented by means of low-level features, but a signi£cant degree of abstraction is achieved if an appropriate representation of the data is associated to a reliable classi£cation procedure. Support Vector Machines (SVMs) [1] gained in the last decade an increasing attention for their good generalization ability even when the dimension of the input space is very high — as in the case of images. An important aspect of statistical learning approaches like Support Vector Machines is which kernel to use for which problem. A number of general purpose kernels is available in the literature but appropriate kernel functions can lead to a substantial increase in the generalization ability of the learning system developed.

In this paper we describe how SVMs, equipped with an appropriate kernel function — the histogram intersection kernel — allow us to obtain very good results on color-based image classi£cation. Histogram intersection [2] is a similarity measure well known in the computer vision literature as an effective indexing technique for color-based recognition. We show that this function is a Mercer's kernel, and thus can be used as a kernel function for SVMs [3].

In the literature, the main contribution on color kernels comes from [4], where a family of functions better suited than Gaussian kernels for dealing with image histograms is studied. In essence the kernels described in [4] ensure heavier tails than Gaussian kernels in the attempt to contrast the well known phenomenon of diagonally dominant kernel matrices in the case of high dimensional inputs. We present experimental results that show how histogram intersection kernel is superior to general purpose kernels and performs very well compared with Heavy-tailed RBF kernels.

The paper is organized as follows. In Section 2 we provide a brief introduction to SVMs; histogram intersection is introduced in Section 3, in Section 4 we show why histogram intersection is a Mercer's kernel. In Section 5 we describe the results of experiments on two applications: indoor-outdoor classi£cation and dominant color perception. We draw the conclusions of our work in Section 6.

2. SUPPORT VECTOR MACHINES

Following [5], many problems of statistical learning [1] can be cast in an optimization framework in which the goal is to determine a function minimizing a functional I of the form

$$I[f] = \frac{1}{\ell} \sum_{i=1}^{\ell} V(f(\mathbf{x}_i), y_i) + \lambda \|f\|_K^2. \qquad (1)$$

The ℓ pairs $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_\ell, y_\ell)\}$, the examples, are *i.i.d.* random variables drawn from the space $X \times Y$ according to some £xed but unknown probability distribution, V is a *loss* function measuring the £t of the function f to the data, $\|\cdot\|_K$ the norm of f induced by a certain function K, named *kernel*, controlling the smoothness – or capacity – of f, and $\lambda > 0$ a trade-off parameter. For several choices of the loss function V, the minimizer of the functional in (1) takes the general form

$$\sum_{i=1}^{\ell} \alpha_i K(\mathbf{x}, \mathbf{x}_i), \qquad (2)$$

where the coef£cients α_i depend on the examples. The mathematical requirements on K must ensure the convexity of (1) and hence the uniqueness of the minimizer (2). This is guaranteed by requiring the positive de£niteness of the function K. A theorem of functional analysis due to Mercer allows us to write a positive de£nite function as an expansion of certain functions $\phi_k(\mathbf{x})$ $k = 1, ..., N$, with N possibly in£nite, or

$$K(\mathbf{x}, \mathbf{x}') = \sum_{k=1}^{N} \phi_k(\mathbf{x})\phi_k(\mathbf{x}'). \tag{3}$$

A positive de£nite function is also called a *Mercer's kernel*, or simply a *kernel* [6, 7]. SVMs for classi£cation [1] correspond to choices of V like $V(f(\mathbf{x}_i), y_i) = |1 - y_i f(\mathbf{x}_i)|_+$, with $|t|_+ = t$ if $t \geq 0$, and 0 otherwise, and lead to a convex QP problem with linear constraints in which many of the α_i vanish. The points $\mathbf{x}_i$ for which $\alpha_i \neq 0$ are termed *support vectors* and are the only examples needed to determine the solution (2).

2.1. Binary classi£cation

In the case of binary classi£cation we have $y_i \in \{-1, 1\}$ for $i = 1, ..., \ell$, and the dual optimization problem can be written as [1]

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & \sum_{i=1}^{\ell} |\alpha_i| - \sum_{i,j=1}^{\ell} \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (4) \\ \text{subject to} \quad & \sum_{i=1}^{\ell} \alpha_i = 0, \quad 0 \leq (y_i \alpha_i) \leq C \end{aligned}$$

A new point is classi£ed according to the sign of the expression $\sum_{i=1}^{\ell} \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b$, where the coef£cient b can be determined from the Kuhn-Tucker conditions.

2.2. Novelty detection

In the case of novelty detection or one class classi£cation [8], for all training points we have $y_i = 1$ and the optimization problem reduces to

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & \sum_{i=1}^{\ell} \alpha_i K(\mathbf{x}_i, \mathbf{x}_i) - \sum_{i,j=1}^{\ell} \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (5) \\ \text{subject to} \quad & \sum_{i=1}^{\ell} \alpha_i = 1 \quad 0 \leq \alpha_i \leq C. \end{aligned}$$

If $K(\mathbf{x}, \mathbf{x})$ is constant over the domain X, a novelty is detected if the inequality $\sum_{i=1}^{\ell} \alpha_i K(\mathbf{x}, \mathbf{x}_i) \geq \tau$ is violated for some £xed value of the threshold parameter $\tau > 0$.

3. HISTOGRAM INTERSECTION

Possibly the simplest way to represent color information in images is provided by histograms. Different color spaces (RGB, CMYK, HSV, etc.) give rise to different representations of the images that enhance or attenuate certain color features. In spite of their simplicity, color histograms are very ef£cient for many practical tasks, as they are stable to occlusions and to change of view [2]: in building color histograms all information regarding spatial features and spatial correlation is discarded. This leads to invariance to rotation and tolerance to translation, but, on the other hand, limits their use if spatial correlation is an important cue.

There are many examples in nature where color or light correlate with the identity of an object [2] or of a class of objects; for this reason histograms are a suitable representation of images for many indexing and categorization tasks. *Histogram intersection* is a technique proposed in [2] for color indexing with application to object recognition. From the results reported in [2] and in subsequent works we know that histogram intersection is an effective representation which makes it possible to build reasonably effective color-based recognition systems. Color histogram intersection K_{int} measures the degree of similarity between two color histograms. It is well suited to deal with scale changes and can be successfully used even in the presence of non-segmented objects. Histogram intersection can be de£ned as follows. We denote with A and B the histograms of images A_{im} and B_{im}. Both histograms consist of m bins, and the i-th bin for $i = 1, ..., m$ is denoted with a_i and b_i respectively. Let us assume that A_{im} and B_{im} have the same size (N pixels); by construction we have $\sum_{i=1}^{m} a_i = N$ and $\sum_{i=1}^{m} b_i = N$. Then,

$$K_{int}(A, B) = \sum_{i=1}^{m} \min\{a_i, b_i\}. \tag{6}$$

4. IS HISTOGRAM INTERSECTION A KERNEL?

In this section we demonstrate that histogram intersection K_{int} *is* a Mercer's kernel. We prove it by showing that histogram intersection is an inner product in a suitable feature space. As stated in the previous section, let A and B denote the histograms of images A_{im} and B_{im}. We now represent A with an $N \times m$-dimensional binary vector $\mathbf{A}$ de£ned as

$$\begin{aligned} \mathbf{A} \quad = \quad & (\overbrace{1, 1, ..., 1}^{a_1}, \underbrace{0, ..., 0}_{N-a_1}, \quad (7) \\ & \overbrace{1, 1, ..., 1}^{a_2}, \underbrace{0, 0, ..., 0}_{N-a_2}, ..., \overbrace{1, 1, ..., 1}^{a_m}, \underbrace{0, 0, ..., 0}_{N-a_m}), \end{aligned}$$

and similarly B with $\mathbf{B}$. Notice the redundant representation of the information content of histogram A in the binary vector $\mathbf{A}$ of (7). The histogram intersection $K_{int}(A, B)$ in (6) can be readily seen to be equal to the standard inner product between the two corresponding vectors $\mathbf{A}$ and $\mathbf{B}$:

$$K_{int}(A, B) = \mathbf{A} \cdot \mathbf{B}. \tag{8}$$

We thus have that histogram intersection is a Mercer's kernel and Eq. 8 describes explicitly Eq. 3. Two comments are in order. First, the generalization of this result to higher dimensions like 3-D color space representations is straightforward. Second, in the case of images of different size all what is needed is to standardize the bins' number, normalize the histogram areas, and repeat the same construction of above on the resulting histograms.

5. EXPERIMENTS

In the above section we showed that histogram intersection has the mathematical properties required for it to be a suitable kernel function for SVMs. Here we discuss its effectiveness, as a similarity measure, in practical applications of image classi£cation. We compare the performance of histogram intersection with classical polynomial and Gaussian RBF kernels, and also with representatives of the family of non-Gaussian RBF kernels [4]. This general class of kernels can be written as $K(\mathbf{x}, \mathbf{z}) = \exp\left(-\frac{d(\mathbf{x},\mathbf{z})}{\sigma^2}\right)$, where d is a suitable distance function. In [4] the authors discuss possible choices of d, investigating among functions extensively used for histogram comparison (such as χ^2), and among Heavy-tailed RBF kernels (that lead to Mercer's kernels if $b \leq 2$ [1]):

$$d_{a,b}(\mathbf{x}, \mathbf{z}) = \sum_i |x_i^a - z_i^a|^b. \qquad (9)$$

5.1. Indoor-outdoor classi£cation

Deciding whether an image is an indoor or an outdoor view is a typical example of understanding the "semantic" of the image content [9]. Its practical use ranges from scene understanding, to automatic image adjustment, to intelligent £lm development. In a £rst series of experiments we trained a binary SVM with a set of 300 images acquired by 6 digital cameras differing in quality and manufacture. Images were acquired under different light conditions and different camera settings by 10 people. The results of these experiments were promising: we obtained, for various bins' choices, recognition rates around 90% for a RGB representation and rates above 92% for a HSV representation. We moved to a dataset of images downloaded from the Web to prevent the results to be biased by the limited number of cameras (since different optics in¤uence color formation in different ways) and users. Again, we trained a number of SVMs for binary classi£cation on a set of 300 indoor and 300 outdoor images (downloaded from http://www.benchathlon.net/ http://www.cs.washington.edu/research/imagedatabase/; see Fig. 1) of various size – typically of $10^4 \div 10^5$ pixels. Each image was used to construct a color histogram in the HSV color space consisting of $15 \times 15 \times 15$ bins. The number of bins was chosen as a good compromise between computational cost and adequacy of the description. The recognition rates on a test set of 123 indoor images and 260 outdoor images, for various kernels, are shown in Table 1. The recognition rate of the color histogram intersection kernel, slightly above 93%, appears to be $3\% \div 4\%$ better than the best general purpose kernels (see Table 1). The results are comparable to the ones obtained with Heavy-tailed RBF, for suitably chosen parameters: with a Laplacian RBF kernel and a suitable σ we obtained a recognition rate of about 92%. The choice of a and b that produced the best results for image categorization in [4] (i.e., $a = 0.25$, $b = 1$) does not give good results on the indoor-outdoor problem on our database (the recognition rate obtained is 86.9%). The main point, though, is not that histogram intersection performs at least 1% better than other kernels, but that it does not require the process of seeking the optimal combination of parameters, allowing us to design a system which is easier to tune. Our results can be compared to those reported in [9] in which a fairly complex classi£cation procedure was proposed and tested on a different database reaching a recognition rate slightly better than 90%. Overall, the good recognition rates obtained by all kernels are presumably due to the fact that the color histogram representation captures an essential aspect of color appearance for this problem.

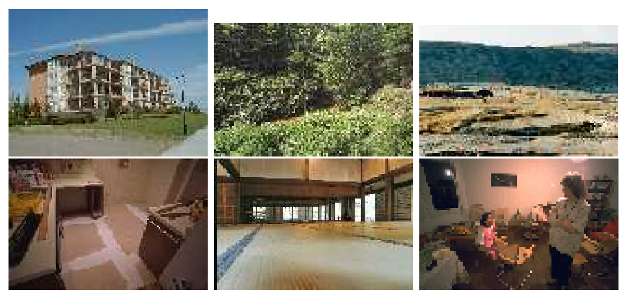

Fig. 1. Some of the images of indoor and outdoor scenes, downloaded from the Web, used in the experiments.

5.2. Perceiving dominant colors

This section describes work in progress. We are studying whether it may be possible to train a learning machine to understand if an image would give to a human observer the impression of a dominant color. It is not clear how this issue of perceived color could be addressed using a straightforward algorithmic approach and the nature of the problem may in fact be rather different. We are carrying out our experiments on the Artchive database (http://www.artchive.com), that contains thousands of images of paintings. The idea is to train a machine to "simulate" the perception of color of one or more people. We obtained preliminary results which are very promising by training a system on a dataset manually extracted by two different people. Here we present the

Kernel type	**r.r.** (%)
histogram intersection	93.1
linear	88.9
2-nd deg polynomial	89.2
3-rd deg polynomial	89.4
4-th deg polynomial	88.1
Gaussian RBF kernel ($\sigma = 0.3$)	86.5
Gaussian RBF kernel ($\sigma = 0.5$)	87.8
Gaussian RBF kernel ($\sigma = 0.7$)	89.1
Gaussian RBF kernel ($\sigma = 0.9$)	88.9
Laplacian RBF kernel ($\sigma = 1.0$)	92.1
Laplacian RBF kernel ($\sigma = 0.8$)	88.1
Laplacian RBF kernel ($\sigma = 1.2$)	90.0
Sublinear RBF kernel ($\sigma = 4.0$)	88.7

Table 1. Recognition rates (*r.r.*) for SVMs with different kernels on the indoor-outdoor classi£cation problem. Laplacian RBF corresponds to the choice of a=1 and b=1 (Eq. 9), sublinear RBF are such that a=1 and b=0.5.

results obtained on representing the perception of a green dominant on images of paintings. To this purpose we trained a one class SV-based classi£er on a dataset of 51 positive examples. We tested the system on a test set of 53 positive examples and 164 negative examples — positive examples test the robustness to false negatives, negative examples to false positives. The results obtained, in terms of *equal error rate*, i.e., the error obtained for the best compromise between false positives and false negatives, are summarized in Table 2. The performance obtained with histogram intersection kernel is well above standard kernels and comparable to a Heavy-tailed kernel with suitable parameters.

Kernel type	**HSV** (%)	**RGB** (%)
histogram intersection	85	86
2-nd deg polynomial	74	70
3-rd deg polynomial	69	70
Gaussian RBF kernel ($\sigma = 1.0$)	71	68
Gaussian RBF kernel ($\sigma = 1.5$)	70	71
Laplacian RBF kernel ($\sigma = 2$)	85	85
RBF kernel a=1 b=1.5 ($\sigma = 2$)	72	72

Table 2. E.E.R. for SVMs with different kernels, and for RBG and HSV representations, on the problem of "simulating" the perception of a dominant color.

6. CONCLUSION

We presented a study on image classi£cation based on color global features. We showed that histogram intersection ful£lls the mathematical requirements for it to be used as a kernel for SVMs. We also showed that it performs well, compared with standard kernels and with other state-of-the-art color kernels, and it has the nice property of being easy to tune, since it depends on fewer parameters. We presented experimental results on two applications: indoor-outdoor classi£cation and dominant color perception.

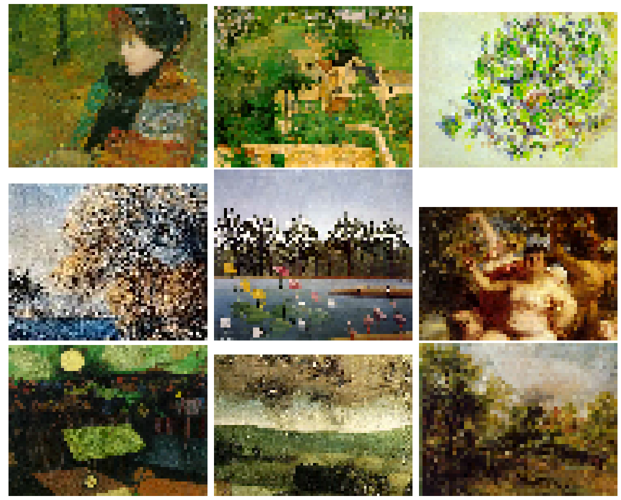

Fig. 2. Images used in the experiments on the perception of dominant green. From top to bottom: 3 examples of training data, and examples used for testing (3 neg. and 3 pos.).

7. REFERENCES

[1] V. Vapnik, *Statistical learning theory*, John Wiley and sons, New York, 1998.

[2] M. J. Swain and D. H. Ballard, "Color indexing," *IJCV*, vol. 7, no. 1, pp. 11–32, 1991.

[3] A. Barla, E. Franceschi, F. Odone, and A. Verri, "Image kernels," in *Int. Work. on Pattern Recognition with SVM, ICPR 2002, LNCS 2388*, 2002.

[4] I. Chapelle, P. Haffner, and V. Vapnik, "SVMs for histogram-based image classi£cation," *IEEE Transactions on Neural Networks, special issue on Support Vectors*, 1999.

[5] T. Evgeniou, M. Pontil, and T. Poggio, "Regularization networks and support vector machines," *Advances in Computational Mathematics*, vol. 13, pp. 1–50, 2000.

[6] R. Courant and D. Hilbert, *Methods of Mathematical Physics, Vol. 2*, Interscience, 1962.

[7] T. Poggio, S. Mukherjee, R. Rifkin, A. Rakhlin, and A. Verri, "b," 2002.

[8] C. Campbell and K. P. Bennett, "A linear programming approach to novelty detection," *Advances in Neural Information Processing Sys.*, vol. 13, 2001.

[9] M. Szummer and R. W. Picard, "Indoor-outdoor image classi£cation," in *IEEE Intl Workshop on Content-based Access of Image and Video Databases*, 1998.