

Hausdorff Kernel for 3D Object Acquisition and Detection

Annalisa Barla, Francesca Odone, and Alessandro Verri

INFM - DISI, Università di Genova
Via Dodecaneso 35, 16146 Genova, Italy
{barla,odone,verri}@disi.unige.it

Abstract. Learning one class at a time can be seen as an effective solution to classification problems in which only the positive examples are easily identifiable. A kernel method to accomplish this goal consists of a representation stage - which computes the smallest sphere in feature space enclosing the positive examples - and a classification stage - which uses the obtained sphere as a decision surface to determine the positivity of new examples. In this paper we describe a kernel well suited to represent, identify, and recognize 3D objects from unconstrained images. The kernel we introduce, based on Hausdorff distance, is tailored to deal with grey-level image matching. The effectiveness of the proposed method is demonstrated on several data sets of faces and objects of artistic relevance, like statues.

1 Introduction

In the *learning from examples* paradigm the goal of many detection and classification problems of computer vision – like object detection and recognition (see [9,10,11,12,13,20] for example) – is to distinguish between positive and negative examples. While positive examples are usually defined as images or portion of images containing the object of interest, negative examples are comparatively much less expensive to collect but somewhat ill defined and difficult to characterize. Problems for which only positive examples are easily identifiable are sometimes referred to as novelty detection problems. In this paper we study a kernel method for learning one class at a time in a multiclass classification problem.

Kernel methods, which gained an increasing amount of attention in the last years after the influential work of Vapnik [17,18], reduce a learning problem of classification or regression to a multivariate function approximation problem in which the solution is found as a linear combination of certain positive definite functions named kernels, centered at the examples [3,4,19]. If the examples belong to only one class, the idea is that of determining the spatial support of the available data by finding the smallest sphere in feature space enclosing the examples [1]. The feature mapping, or the choice of the norm, plays here a crucial role.

In this paper, we introduce a kernel derived from an image matching technique based on the notion of Hausdorff distance [7], well suited to capture image similarities, while preserving meaningful image differences. The proposed kernel is tolerant to small amount of local deformations and scale changes, does not require accurate image registration or segmentation, and is well suited to deal with occlusion issues.

We present experiments which show that this kernel outperforms the linear kernel (effectively corresponding to a standard template matching technique) and polynomial kernels in the representation and the identification of 3D objects. The efficacy of the method is assessed on databases of faces and 3D objects of artistic interest. All the images used in the reported experiments are available for download at <ftp://ftp.disi.unige.it/person/OdoneF/3Dobjects>.

In summary, the major aim of this paper is to evaluate the appropriateness of Hausdorff-like measures for engineering kernels well suited to deal with image related problems. A secondary objective is to assess the potential of kernel methods in computer vision, in the case in which a relatively small number of examples of only one class is available. The paper is organized as follows. The kernel method used in this paper, suggested by Vapnik [17] and developed in [1], is summarized in Section 2. Section 3 introduces and discusses the Hausdorff kernel for images. The experiments are reported in Section 4, Section 5 is left to conclusions.

2 Kernel-Based Approach to Learn One Class at a Time

In this section we review the method described in [1] which shares strong similarities with Support Vector Machines (SVMs) [17,18] for binary classification. The main idea behind this approach is to find the sphere in feature space of minimum radius which contains most of the data of the training set. The possible presence of outliers is countered by using slack variables ξ_i which allow for data points outside the sphere. This approach was first suggested by Vapnik [17] and interpreted and used as a novelty detector in [15] and [16]. If R is the sphere radius and \mathbf{x}_0 the sphere center, the primal minimization problem can be written as

$$\begin{aligned} \min_{R, \mathbf{x}_0, \xi_i} \quad & R^2 + C \sum_{i=1}^{\ell} \xi_i \\ \text{subject to} \quad & (\mathbf{x}_i - \mathbf{x}_0)^2 \leq R^2 + \xi_i \quad \text{and} \quad \xi_i \geq 0, \quad i = 1, \dots, \ell. \end{aligned} \quad (1)$$

with $\mathbf{x}_1, \dots, \mathbf{x}_\ell$ the input data, $\xi_i \geq 0$ and C a regularization parameter. The dual formulation requires the solution of the QP problem

$$\begin{aligned} \max_{\alpha_i} \quad & - \sum_{i=1}^{\ell} \alpha_i \mathbf{x}_i \cdot \mathbf{x}_i + \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \alpha_i \alpha_j \mathbf{x}_i \cdot \mathbf{x}_j \\ \text{subject to} \quad & \sum_{i=1}^{\ell} \alpha_i = 1, \quad \text{and} \quad 0 \leq \alpha_i \leq C. \end{aligned} \quad (2)$$

As in the case of SVMs for binary classification, the objective function is quadratic, the Hessian positive semi-definite, the inequality constraints are box constraints. The two main differences are the form of the linear term in the objective function and the equality constraint (the Lagrange Multipliers sum to 1 instead than 0). Like in the case of SVMs, the training points for which $\alpha_i > 0$ are the *support vectors* for this learning problem.

The sphere center \mathbf{x}_0 is found as a weighted sum of the examples as $\mathbf{x}_0 = \sum \alpha_i \mathbf{x}_i$ while the radius R can be determined from the Kuhn-Tucker condition associated to any training point \mathbf{x}_i for which $0 < \alpha_i < C$ as

$$R^2 = (\mathbf{x}_i - \mathbf{x}_0)^2.$$

The distance between a point and the center can be computed with the following equation

$$\begin{aligned} d^2(\mathbf{x}) &= (\mathbf{x} - \mathbf{x}_0)^2 = \\ &= \mathbf{x} \cdot \mathbf{x} - 2\mathbf{x} \cdot \mathbf{x}_0 + \mathbf{x}_0 \cdot \mathbf{x}_0 = \\ &= \mathbf{x} \cdot \mathbf{x} - 2 \sum_{i=1}^{\ell} \alpha_i \mathbf{x} \cdot \mathbf{x}_i + \sum_{i,j=1}^{\ell} \alpha_i \alpha_j \mathbf{x}_i \cdot \mathbf{x}_j \end{aligned}$$

In full analogy to the SVM case, one can introduce a *kernel function* K [18], and solve the problem

$$\begin{aligned} \max_{\alpha_i} \quad & - \sum_{i=1}^{\ell} \alpha_i K(\mathbf{x}_i, \mathbf{x}_i) + \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) \\ \text{subject to} \quad & \sum_{i=1}^{\ell} \alpha_i = 1 \quad \text{and} \quad 0 \leq \alpha_i \leq C, \quad i = 1, \dots, \ell. \end{aligned} \quad (3)$$

A kernel function K is a function satisfying certain mathematical constraints [2, 18] and implicitly defining a mapping ϕ from the input space to the feature space – space in which the inner product between the feature points $\phi(\mathbf{x}_i)$ and $\phi(\mathbf{x}_j)$ is $K(\mathbf{x}_i, \mathbf{x}_j)$. The constraints on α_i define the *feasible* region of the QP problem.

In this case the sphere center in feature space cannot be computed explicitly, but the distance $d_K(\mathbf{x})$ between the sphere center and a point \mathbf{x} can be written as

$$d_K^2(\mathbf{x}) = K(\mathbf{x}, \mathbf{x}) - 2 \sum_{i=1}^{\ell} \alpha_i K(\mathbf{x}, \mathbf{x}_i) + \sum_{i,j=1}^{\ell} \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j).$$

As in the linear case, the radius R_K can be determined from the Kuhn-Tucker conditions associated to a support vector \mathbf{x}_i for which $0 < \alpha_i < C$.

3 The Hausdorff Kernel

In this section we first describe a similarity measure for images inspired by the notion of Hausdorff distance. Then, we determine the conditions under which this measure defines a legitimate kernel function.

3.1 Hausdorff Distances

Given two finite point sets A and B (both subsets of \mathbb{R}^N), the *directed Hausdorff distance* h , can be written as

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|.$$

Clearly the directed Hausdorff distance is not symmetric and thus not a “true” distance, but it is very useful to measure the degree of mismatch of one set with respect to another. To obtain a distance in the mathematical sense, symmetry can be restored by taking the maximum between $h(A, B)$ and $h(B, A)$. This brings to the definition of *Hausdorff distance*, that is,

$$H(A, B) = \max\{h(A, B), h(B, A)\}.$$

A way to gain intuition on Hausdorff measures which is very important in relation to the similarity method we are about to define, is to think in terms of set inclusion. Let B_ρ be the set obtained by replacing each point of B with a disk of radius ρ , and taking the union of all of these disks; effectively, B_ρ is obtained by dilating B by ρ . Then the following holds:

Proposition The directed Hausdorff distance $h(A, B)$ is not greater than ρ if and only if $A \subseteq B_\rho$.

This follows easily from the fact that, in order for every point of A to be within distance ρ from some points of B , A must be contained in B_ρ .

3.2 Hausdorff-Based Measure of Similarity between Grey-Level Images

Suppose to have two grey-level images, I_1 and I_2 , of which we want to compute the degree of similarity; ideally, we would like to use this measure as a basis to decide whether the two images contain the same object, maybe represented in two slightly different views, or under different illumination conditions. In order to allow for grey level changes within a fixed interval or small local transformations (for instance small scale variations or affine transformations), a possibility is to evaluate the following function [7]

$$k(I_1, I_2) = \sum_p \theta(\epsilon - \min_{q \in N_p} |I_1[p] - I_2[q]|) \quad (4)$$

where θ is the unit step function. The function k counts the number of pixels p in I_1 which are within a distance ϵ (on the grey levels) from at least one pixel q of I_2 in the neighborhood N_p of p . Unless N_p coincides with p , k is not symmetric, but symmetry can be restored by taking the max as for the Hausdorff Distance, or the average

$$K = \frac{1}{2}[k(I_1, I_2) + k(I_2, I_1)] \quad (5)$$

Equation (4), can be interpreted in terms of set dilation and inclusion, leading to an efficient implementation [7] which can be summarized in three steps.

1. Expand the two images I_1 and I_2 into 3D binary matrices \mathcal{I}_1 and \mathcal{I}_2 , the third dimension being the grey value:

$$\mathcal{I}_1(i, j, g) = \begin{cases} 1 & \text{if } I_1(i, j) = g; \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

2. Dilate both matrices by growing their nonzero entries by a fixed amount ϵ in the grey value dimension, ϵ_r and ϵ_c (the size of the neighbourhood N_p) in the space dimensions. Let \mathcal{D}_1 and \mathcal{D}_2 be the resulting 3D dilated binary matrices. This dilation varies according to the degrees of similarity required and the transformations allowed.
3. Compute the size of the intersections between \mathcal{I}_1 and \mathcal{D}_2 , and \mathcal{I}_2 and \mathcal{D}_1 and take the average of the two values obtaining $K(I_1, I_2)$.

3.3 Relationship with the Hausdorff Distance

The similarity measure k is closely related to the directed Hausdorff distance h : computing k is equivalent to fix a maximum distance ρ_{max} (by choosing ϵ and N_p) allowed between two sets, and see if the sets we are comparing, or subsets of them, are within that distance. In particular, if the dilation is isotropic in an appropriate metric, and $k(I_1, I_2)$ takes the maximum value s (which means that there is a total inclusion of one image in the dilation of the other), then $h(I_1, I_2) \leq \rho_{max}$.

In general, if $k(I_1, I_2) = m < s$ we say that the m -partial directed Hausdorff distance [6] is not greater than ρ_{max} , which means, loosely speaking, that a subset of I_1 , of cardinality m is within a distance ρ_{max} from I_2 .

3.4 Is K a True Kernel?

A sufficient condition for a function K to be used as a kernel is the positive definiteness: given a set $X \subseteq \mathbb{R}^n$, a function $K : X \times X \rightarrow \mathbb{R}$ is *positive definite* if for all integers n and all $\mathbf{x}_1, \dots, \mathbf{x}_n \in X$, and $\alpha_1, \dots, \alpha_n \in \mathbb{R}$,

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) \geq 0. \quad (7)$$

In the case of interest the inequality (7) is always satisfied because, by definition, $K(\mathbf{x}_i, \mathbf{x}_j) \geq 0$ for each \mathbf{x}_i and \mathbf{x}_j , and also in the feasible region of the optimization problem (3) all the α_i are non-negative. In the case of binary classification, case in which the α_i can also be negative, the proposed function is not a kernel, unless the dilation is appropriately redefined (see [8] for details).

4 Experiments

In this section we present results obtained on several image data sets, acquired from multiple views: a group of data sets of faces for face recognition, and group

of data set of 3D objects of artistical interest acquired in San Lorenzo Cathedral. The two problems are closely related but distinct. As we will see in the case of the latter, the notion of object is blurred with the notion of scene (the background being as important as the foreground), while in face recognition this is not the case.

In our approach we make a minimal use of preprocessing. In particular, we do not compute accurate registration, since our similarity measure takes care of spatial misalignments. Also, since our mapping in feature space allows for some degree of deformation both in the grey-levels and in space, the effects of small illumination and pose changes are attenuated. Finally, we exploit full 3D information on the object by acquiring a training set which includes frontal and lateral views.

In the remainder of this section we will first describe the recognition system, and then results of the applications on the two different families of data sets. All the images used have been resized to 72×57 pixels, therefore we work in a space of more than 4000 dimensions. The results presented in this section have been obtained by applying a dilation of 1 on the spatial direction and a dilation of 3 on the grey-levels.

4.1 The Recognition System

Given a training set of images of the object of interest, we estimate the smallest sphere containing the training data in the feature space implicitly defined by the kernel K as described in Section 3. After training, a new image \mathbf{x} is classified as a positive example if

$$K(\mathbf{x}, \mathbf{x}) - 2 \sum_{j=1}^{\ell} \alpha_j K(\mathbf{x}, \mathbf{x}_j) + \sum_{j,k=1}^{\ell} \alpha_j \alpha_k K(\mathbf{x}_j, \mathbf{x}_k) = d_K^2(\mathbf{x}) \leq t. \quad (8)$$

where $t \geq 0$ is a threshold typically in the range of the square of radius R_K . It is interesting to remark that while in the case of the linear kernel and polynomial kernels all the terms in the l.h.s. of (8) need to be computed for each point, for the Hausdorff kernel the only term which depends on point \mathbf{x} is the second, the other two being constant. In this case the inequality $d_K^2(\mathbf{x}) \leq t$ can be rewritten as

$$\sum \alpha_i K(\mathbf{x}, \mathbf{x}_i) \geq \tau, \quad (9)$$

for some suitably defined threshold τ .

4.2 Application to Face Identification

First Data Set

In the first set of images, we acquired both training and test data in the same session. We collected four sets of images (frontal and rotated views), one for each of four subjects, for a total of 353 images, samples of which are shown in Figure 1. To test the system we used 188 images of ten different subjects, including test

images of the four people used to acquire the training images (see the examples in Figure 2).

All the images were acquired in the same location and thus have a similar background. No background elimination was performed since the face occupies a substantial part of the image – about three quarters of the total image area, — but this implies that even images of different people have, on average, one fourth of the pixels which match. This is shown in Figure 3: the white pixels in the rightmost image are points of the first image close in the Hausdorff sense to the middle image — they mainly correspond to background pixels. Notice that the two binary images are similar even though one of them has been computed by comparing two images of the same subject, while the other one by comparing two images of different subjects. This makes it clear that in this setting one or a few images are not sufficient for object characterization. The performance of



Fig. 1. Two training images for each of the four subjects

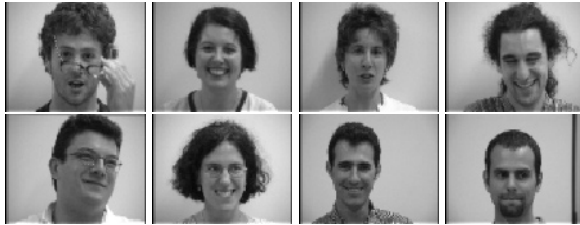


Fig. 2. Examples of test images

the Hausdorff kernel have been compared with a linear kernel, on the same set of examples. The choice of the linear kernel is due to the fact that it can be proved to be similar to correlation techniques (sum of squared differences and cross correlation) widely used for the similarity evaluation between grey level images.

The results of this comparison are shown as Receiver Operating Characteristic (ROC) curves [5]. Each point of an ROC curve represents a pair of false-alarms and hit-rate of the system, for a different threshold t (Equation (8)). The system efficiency can be evaluated by the growth rate of its ROC curve:

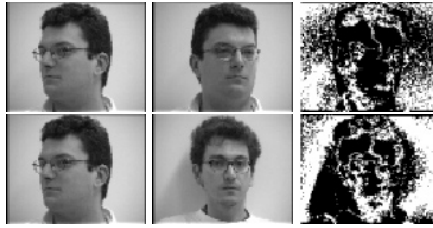


Fig. 3. Spatial support of the Hausdorff distance. Both rows: the white pixels in the rightmost image show the locations of the leftmost image which are close, in the Hausdorff sense, to the middle image.

for a given false-alarm rate, the better system will be the one with the higher hit probability. The overall performance of a system can be measured by the area under the curve. Figure 4 shows the ROC curves of a linear and Hausdorff kernels for all the four face recognition problems. The curve obtained with the Hausdorff kernel is always above the one of the linear kernel, showing superior performance. The linear kernel does not appear to be suitable for the task: in all the four cases, to obtain a hit rate of the 90% with the linear kernel one should accept more than 50% false positives. The Hausdorff kernel ROC curve, instead, increases rapidly and shows good properties of sensitivity and specificity. We tested the robustness of the recognition system by adding difficult positives to one of the test sets (see Figure 5). The corresponding ROC curve graph is the one in the lower left corner of Figure 4.

In a second series of experiments on the same training sets, we estimate the system performance in the leave-one-out mode. We trained the system on $\ell - 1$ examples, and tested it on the one left out, for all possible choices of $\ell - 1$ examples. Figure 6 shows samples from one of the training sets. Even if the set contains a number of images which are difficult to classify, only 19% were found to lie outside the sphere of minimum radius, but all of them within a distance less than 3% of the estimated radius.

Second Data Set

Other sets of face images were acquired on different days and under unconstrained illumination (see examples in Figure 7). A change of scale is immediately noticeable comparing Figure 1 and Figure 7, which therefore necessitated background elimination from the training data: to this purpose, we performed a semi-automatic preprocessing of the training data, exploiting the spatio-temporal continuity between adjacent images of the training set. Indeed, we manually selected a rectangular patch in the first image of each sequence, and then tracked it automatically through the rest of the sequence obtaining the reduced images shown in Figure 8. Figures 9 and 10 show positive and negative test images, respectively, used to test the machine trained with the images represented in Figure 8. The result of the recognition is described by the ROC curve at the top right of Figure 11. For space constraints we did not include examples of images from all the four data sets, but Figure 11 shows the performance of the four systems.

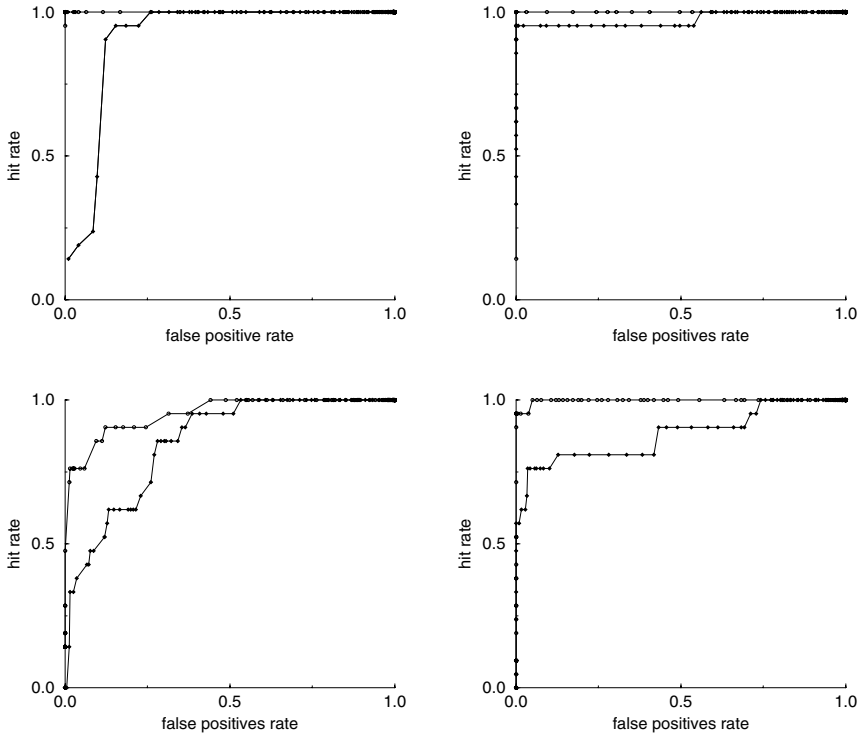


Fig. 4. ROC curves for the four training sets. Comparison between a linear kernel (filled circles) and the Hausdorff kernel (empty circles).



Fig. 5. Examples of the images which affected the ROC curve in the lower left corner of Figure 4



Fig. 6. Examples of the data set used for the leave-one-out experiment

The four different training sets were made of 126, 89, 40, 45 data (corresponding to the curves in clockwise order, from the top left). The size of the test set is specified in the caption of the figure. With this second class of data sets, in most cases the results of the linear kernel are too poor to represent a good comparison, so we also experimented polynomial kernels of various degrees. In the ROCs of Figure 11 we included the results obtained with a polynomial of degree 2, the one which produced the best results. One of the reasons of the failure of classic polynomial kernels may be related to the fact that the data we use are not correlated, i.e., images have not been accurately registered with respect to a common reference. During the acquisition the person was moving and, in different images, the features (eyes, for instance, in the case of faces) are set on different positions.



Fig. 7. One image for each one of the subjects of the new experiments



Fig. 8. Face tracking throughout the sequence. Top rows: original sequences (acquired in two different days); bottom rows: reduced images.



Fig. 9. Samples from positive test sequences, relative to the training set of Figure 8



Fig. 10. Samples from negative test sequences, relative to the the training set of Figure 8

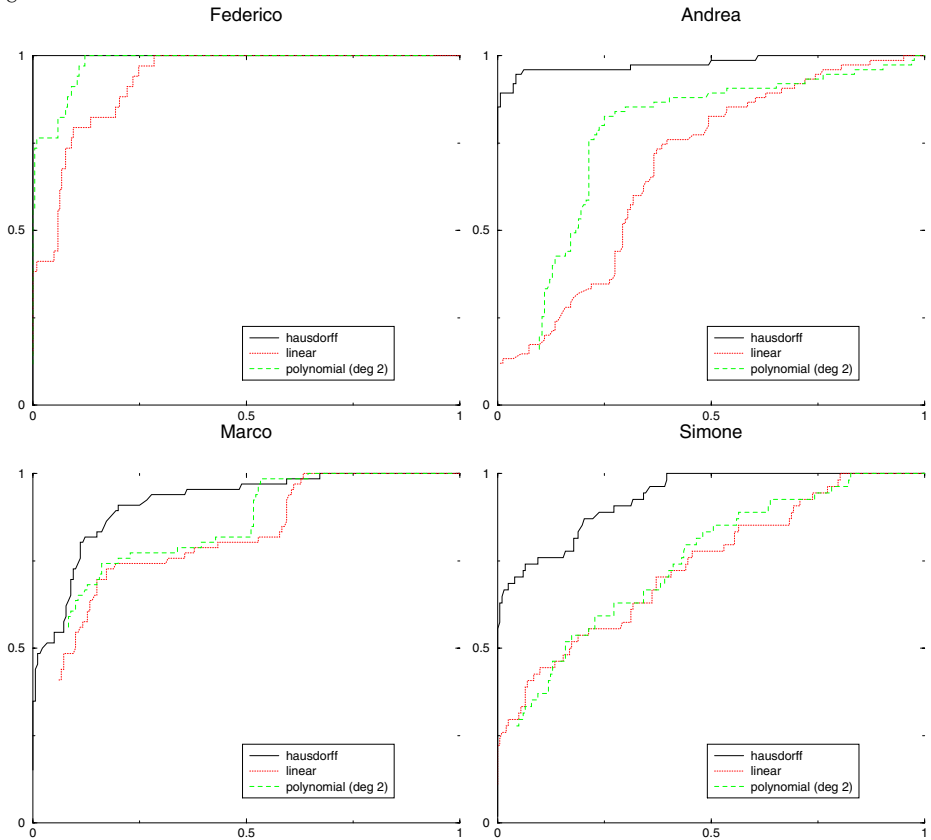


Fig. 11. ROC curves the four subjects of the second data set, comparing the Hausdorff kernel, with linear and polynomial (deg. 2) ones. From top left: Federico (training: 126, test: 102 (positive), 205 (negative)); Andrea (training: 89, test: 228 (positive) 345 (negative)); Marco (training: 40, test: 149 (positive), 343 (negative)); Simone(training: 45, test: 122 (positive), 212 (negative))

4.3 Application to 3D Object Recognition

In this second application we aim to represent and identify 3D objects against a complex background. The data sets, acquired in San Lorenzo Cathedral, are images of marble statues, all located in the same chapel(Cappella di S. Giovanni Battista) and thus acquired under similar illumination conditions: this results in noticeable similarities in the brightness pattern of all the images. In this case

no segmentation is advisable, since the background itself is representative of the object. Here we can safely assume that the statues will not be moved from their usual position.

Figure 12 shows images of two different recognition tasks (two different statues); Figure 13 illustrates negative examples with respect to both statues of Figure 12. Some of the negative examples look alike the positive ones. The performances on the two training sets of Figure 12 are depicted by the ROC curves of Figure 14. We have trained the system with an increasingly small training sets to check the possible degradation of the results: the two curves on the row above have been produced with about 300 training data, the two on the row below with about 95 training data, and a much bigger test set. Notice that the results of our kernel are still very good.



Fig. 12. Samples from two training sets for 2 different statues.

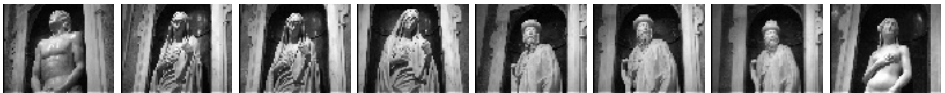


Fig. 13. A few examples of negative examples, with respect to both statues of Figure 12.

5 Conclusion

In this paper we proposed a kernel function derived from the notion of Hausdorff distance for dealing with images of 3D objects. The adopted learning mechanism consists of a kernel method that learns one class at a time, that finds the smallest sphere in feature space containing the training data, and uses it as a decision surface to determine the positivity of new examples. The obtained results indicate that this method, when used with the proposed kernel, appears to be adequate for 3D object modeling, even if the images used are not accurately registered, and thus there is not a precise correspondence between object features in different images. For practical applications the proposed kernel method is suitable for

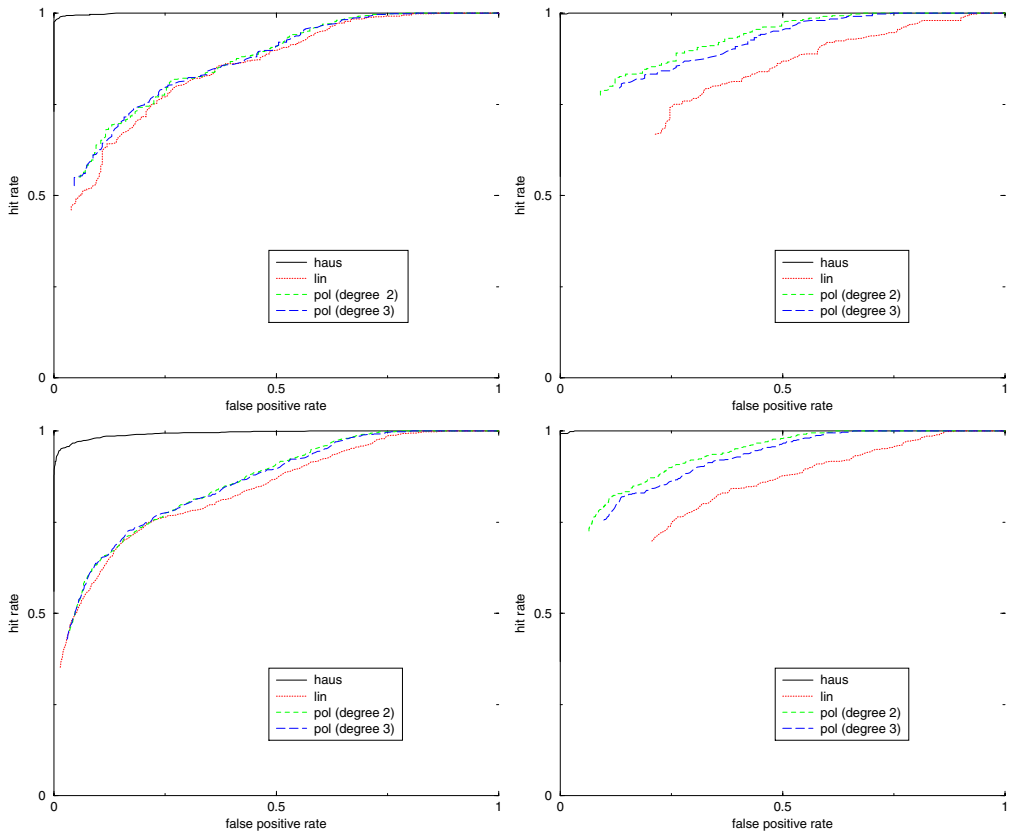


Fig. 14. Effect of the decrease of the training set size: the Hausdorff kernel performs well even with relatively small training sets. Top row: left (training 314, test 628 (pos) 285 (neg)), right (training 226, test 449 (pos) 211 (neg)). Bottom row: left (training 96, test 846 (pos) 3593 (neg)), right (training 97, test 577 (pos) 3500 (neg)).

real-time, provided that reliable prior information is available about object apparent size and rough image location. In the current implementation the average test image is processed in a few seconds on a Pentium III.

Current work on the subject include theoretical studies of the mathematical properties of the Hausdorff kernel including tolerance to occlusions, which, for example, does not apply to Gaussian kernels, and efficient multiscale implementation.

Acknowledgments. We thank Emanuele Franceschi and Fabrizio Smeraldi for many helpful discussions. Francesco Isgrò read the manuscript and made useful remarks. This research was partially supported by the EU Project KerMIT and by the INFM-PRA Project MAIA.

References

1. C. Campbell and K. P. Bennett. A linear programming approach to novelty detection. *Advances in Neural Information Processing Systems*, 13, 2001.
2. N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, 2000.
3. T. Evgeniou, M. Pontil, and T. Poggio. Regularization Networks and Support Vector Machines. *Advances in Computational Mathematics*, 13:1–50, 2000.
4. F. Girosi, M. Jones, and T. Poggio. Regularization Theory and Neural Network Architecture. *Neural Computation*, 7:219–269, 1995.
5. D. M. Green and J. A. Swets. *Signal detection theory and psychophysics*. Reprinted by Krieger, Huntingdon, New York, 1974.
6. D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge. Comparing images using the Hausdorff distance. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 9(15):850–863, 1993.
7. F. Odone, E. Trucco, and A. Verri. General purpose matching of grey level arbitrary images. In G. Sanniti di Baja C. Arcelli, L. P. Cordella, editor, *4th International Workshop on Visual Forms*, Lecture Notes on Computer Science LNCS 2059, pages 573–582. Springer, 2001.
8. F. Odone and A. Verri. Real Time Image Recognition, in *Opto-Mechatronics Systems Handbook*, H. Cho editor; CRC Press. To appear.
9. C. Papageorgiou and T. Poggio. A trainable system for object detection. *International Journal of Computer Vision*, 38:15–33, 2000.
10. M. Pontil and A. Verri. Support Vector Machines for 3D object recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 637–646, 1998.
11. T.D. Rikert, M.J. Jones, and P. Viola. A cluster-based statistical model for object detection. in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1999.
12. H.A. Rowley, S. Baluja, and T. Kanade. Neural Network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:23–38, 1998.
13. H. Schneiderman and T. Kanade. A statistical method for 3-D object detection applied to faces and cars. in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2000.
14. B. Schölkopf, J.C. Platt, J. Shawe-Taylor, A.J. Smola, and R. C. Williamson. Estimating the support of a high-dimensional distribution. Technical Report MSR-TR-99-87, Microsoft Research Corporation, 1999,2000.
15. D. Tax and R. Duin. Data domain description by support vectors. In M. Verleysen, editor, *Proceedings of ESANN99*, pages 251–256. D. Facto Press, 1999.
16. D. Tax, A. Ypma, and R. Duin. Support vector data description applied to machine vibration analysis. In *Proc. of the 5th Annual Conference of the Advanced School for Computing and Imaging*, 1999.
17. V. Vapnik. *The nature of statistical learning theory*. Springer Verlag, Berlin, 1995.
18. V. Vapnik. *Statistical learning theory*. John Wiley and sons, New York, 1998.
19. G. Wahba. *Spline models for observational data*. SIAM, 1990.
20. L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg. A statistical method for 3-D object detection applied to faces and cars. in *Proc. IEEE Int. Conf. on Image Processing*, Wien, 1997.