A Classical Tool Revisited: Object Detection by Statistical Testing

E. Franceschi[1], F. Odone[1], F. Smeraldi[2], and A. Verri[1]

1- INFM – DISI, Università di Genova, Italy
{emafranc,odone,verri}@disi.unige.it

2- Computer Science, Queen Mary, University of London, UK
fabri@dcs.qmul.ac.uk

Abstract: We present a data–driven approach to feature selection and object detection based on hypothesis testing. Starting from positive training examples only, we estimate the probability density of each of a large number of image measurements. A quantitative feature selection criterion inspired by maximum likelihood is then used in conjunction with Spearman's independence rank test to select a maximal subset of discriminative and pairwise independent features. Classification is performed by a sequence of hypothesis tests for the presence of the object. The overall significance level (i.e. the operating point) can be set by controlling the significance level of the individual tests as well as the minimum number of them that a candidate window is required to pass. We report experiments on face detection over the MIT-CBCL database. The image measurements we use for these experiments include grey level values, integral measurements and ranklets. Our results indicate that the method is able to generalize from *positive examples only* and reaches state-of-the-art recognition rates.

Keywords: Object detection, Object recognition, Feature selection, Hypothesis testing, Nonparametrics, Ranklets

1. Introduction

In this paper we discuss a methodology for detecting objects in images based on hypothesis testing. Hypothesis tests appear to be well suited for dealing with detection problems; in particular, they afford a quantitative way to estimate and control the percentage of false negatives by tuning the confidence level. Our work makes use of these classical tools in a new context. We consider the common practical case in which there are enough positive examples to allow reasonable estimates of univariate marginal probability distributions but no clear characterisation of the "reject" class is readily available. One is thus forced to work against the *omnibus* alternative. We show that this situation can be handled by using multiple tests on independent features automatically selected from a large pool of measurements based on their distribution.

In the training stage a very large number of image measurements is collected, and the empirical density of each of them is estimated from the available positive examples. A criterion derived from maximum likelihood is used to identify the most discriminative features. Spearman's independence rank test is then used to further select a maximal subset of pairwise independent features of size N. At run time, a hypothesis test is performed for each feature. The null hypothesis is, in each case, the presence of the object. Detection is achieved if at least M of the N tests are passed. The choice of M is made according to the *overall* confidence level required. The learning process described is efficient in the sense that increasing the number of training samples leads to better estimates of the underlying probability densities without increasing the computational cost at runtime.

Many general–purpose feature selection methods have been proposed (see [9, 1] and references therein). In [8] Adaboost is used to distill a relatively small number of highly descriptive features using information from both positive and negative examples. In a similar context [4] apply a feature selection method based on the analysis of the variance of features to discriminate highly descriptive regions from uniform regions; in [7] a component-based approach is described, where the features are image patches of various sizes and feature selection is performed via mutual information. Neither of these, however, exploit quantitative measures of feature independence, though in [5] some heuristic is used to this purpose. Our work is rooted in classic nonparametric statistical approaches (see [3] for a quite complete overview of this subject), perhaps less popular within the computer vision community than Bayesian or statistical learning techniques, but which appear to be well suited for dealing with detection problems.

We present experimental results on face detection over the images of the MIT-CBCL database, but we strictly view this application as a case study, since our methodology is entirely data driven and does not rely on specific properties of face images. In principle, the porting to a different application is subject only to the availability of a suitable (positive) training set.

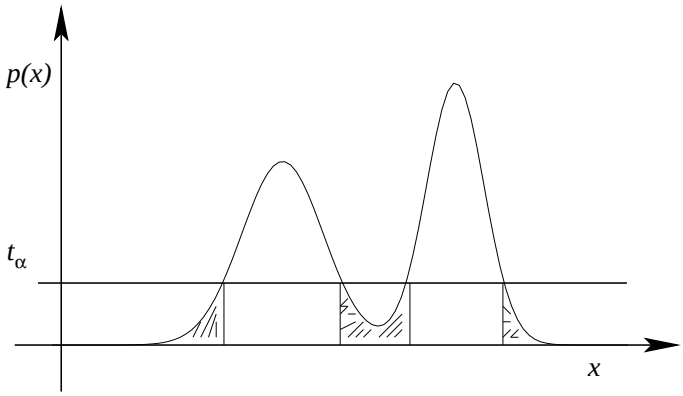

Figure 1: The dashed areas of the distribution $p(x)$ contribute to the "tail" (or the reject region) $t \leq t_\alpha$ of the distribution f defined by Eq. 1.

2. Statistical background

2.1 Testing against the *omnibus* alternative

Hypothesis tests rely on the basic assumptions of knowing the probability distribution of the observable under the null hypothesis *and* a model for the alternative against which the test is run. Possibly the most common choice for an alternative is the shift model, effectively leading to one- or two-sided tests such as, for example, the Student's one-sample *t*-test.

While density estimation is a recognised research topic in itself, the choice of the alternative is generally relegated to the (oftentimes unprincipled) collection of a set of ad–hoc negative examples. In this work, we restrict ourselves to the simple case in which the univariate densities $p(x)$ of our measurements can be satisfactorily approximated by an histogram over the positive training data, and take a more principled stance on the alternative instead.

To this aim, we perform a change of variables and define the probability density function $f(t)$ as

$$\int_0^t f(z)dz = \int_{-\infty}^{+\infty} p(x)U_0(t - p(x))dx \qquad (1)$$

where $U_0(\cdot)$ is the unit step function. For a fixed $t \geq 0$, the integral on the l.h.s. is equal to the probability of the event $\mathsf{D}_t = p^{-1}([0,t])$ (see the dashed area in Fig. 1). We then perform a one-sided test on $f(t)$ — instead than on $p(x)$ — rejecting the null hypothesis for values of t lower than a critical value t_α. Effectively, this implements the maximum likelihood principle by rejecting the null hypothesis if the observable x falls in a region of small probability (see Fig. 1). As usual, the significance level of the test is given $\alpha = \int_0^{t^\alpha} f(t)dt$. Note that by Eq. 1 the tail of f may account for disjoint intervals of $p(x)$ on the x-axis (see again Fig. 1).

2.2 Spearman's independence rank test

An effective way to estimate independence between two observables (in our case, image features) that may have different measurement units is provided by Spearman's independence rank test [3].

Assume we are given n realisations of two random variables, R and S. Let $\pi_R(r_i)$ and $\pi_S(s_i)$ represent the rank of each observation among those of the respective variable.

The Spearman's statistics $\mathcal{D}$ is defined as $\mathcal{D} = \sum_{i=1}^{n}(\pi_R(r_i) - \pi_S(s_i))^2$. The null distribution of $\mathcal{D}$ is obtained under the assumption that for independent variables all rankings occur with probability $1/n!$. For large n a Normal approximation holds, with the tails corresponding to correlated or anti-correlated variables (i.e., equal or opposite rankings). Thus one runs a test against the independence hypothesis with significance α by checking whether $\mathcal{D}$ deviates from its average by more than some critical value d_α.

3. Feature selection

Assume we are given a training set of positive examples only for the object of interest, and a large set of possible image measurements (e.g. grey level values, wavelet coefficients, rank features). In this section we describe a selection procedure that distills a subset of descriptive and independent features for the problem at hand. In the first stage, after estimating the probability distribution of each image measurement from the training set, we select a subset of features according to the notion of saliency defined below.

3.1 Selection of salient features

Considering the type of hypothesis test based on the probability density f of Eq. 1, a natural definition of saliency can be given in terms of t_α. For a fixed significance level α, the image measurement with the cumulative distribution leading to the highest critical value t_α is assigned the maximum saliency. This criterion can be implemented by ranking the features of a given family based on t_α and retaining only a certain fraction or number of the top features.

3.2 Selection of independent features

This second step aims at selecting a subset of independent features out of the salient features identified in the first step. The reason for this is to reduce the number of features without compromising the power of the final test.

This selection is performed by computing the Spearman's statistics for all pairs of features of the same category (grey levels, wavelets, etc). For each category, a graph is built of which the single measurements represent the nodes. Given a threshold $0 < \tau < 1$, two

nodes are joined by an edge if the corresponding features *don't* reject the independence hypothesis with a level of significance lower than τ. Finally, maximally complete subgraphs — or *cliques* — are located in each graph. The nodes of the clique correspond to features that are pairwise independent with confidence greater than $1 - \tau$.

4. Testing for the presence of the object

Detection is achieved by performing a hypothesis test of the type described in Section 2.1 for each image measurement. The null hypothesis is, in each case, the presence of the object.

The idea is to gather evidence for rejecting the null hypothesis — that is, that the image represents the object of interest — by testing the N selected, independent features in a sequence. An object is detected if at least M of the N tests are passed. The overall significance level depends on M as well as on the single tests. In general, we will want to choose a high level of significance for the single tests, so that each of them acts as a "weak classifier". The operating point of the system can then be tuned by varying M.

5. Experiments on face detection

In this section we specialise our method to the case of face detection. We use the CBCL-MIT database for training (feature selection) and validation [1].

5.1 Feature extraction

In the present set up, for each image patch of size 19×19 (the size of each of the 2431 images in the training set) we compute the following collection of features: *(i)* 19×19=361 grey values features (one for each image location), *(ii)* 19 vertical, 19 horizontal, and 37+37 diagonal tomographies, for a total of 112 tomographic features, and *(iii)* 5184 horizontal, 5184 vertical, 5184 diagonal ranklets, for a total of 15,552 ranklet features. Overall this amounts to estimating about 16,000 features.

Tomographies are integral measurements, or averages of image grey values computed along specific directions (at the moment limited to vertical, horizontal, and 45° diagonal). For both grey values and tomographies we first equalise images to attenuate the effect of illumination changes.

Ranklets are a family of orientation selective rank features designed in close analogy with Haar wavelets [6]: in particular, the three orientation channels correspond. Whereas Haar wavelets are a set of filters that act linearly on the intensity values of the image, ranklets are defined in terms of the relative order of pixel intensities and are not affected by monotonic transformations of the grey levels.

[1]The training and (positive) test sets we use have been randomly sampled from the database to ensure they are identically distributed.

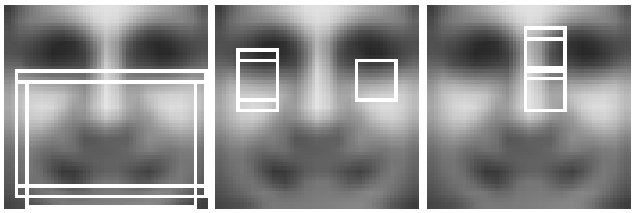

Figure 2: Selected salient features (left to right): the support of the best three diagonal, horizontal and vertical ranklets respectively. The value of the features on the left is distributed around zero: these features capture the symmetry of the face.

The current collection of image features is not exhaustive and can be enriched; we simply regard it as a starting point for validating our method.

5.2 Feature selection

After the selection of the most salient features ranked according to the procedure described in the previous section we are left with 300 grey values, 112 tomographies, 4000 horizontal, 4000 vertical, and 1000 diagonal ranklets. The supports of the more salient nine ranklets are shown in Figure (2).

The subsequent selection of maximal cliques of independent features (with $\tau = 0.5$) left us with ($N =$) 84 grey values, 43 tomographies, 63 horizontal, 75 vertical, and 83 diagonal ranklets.

5.3 Testing for the presence of a face

We validated the face detector on the test set of the MIT-CBCL database, that consists of 472 faces and 23,573 non-faces. Since all images are 19×19 pixels the question is simply whether, or not, an image is a face image.

We first ran our experiments using features from one category only. The results shown in Fig. 3 as ROC curves indicate that grey values and horizontal and vertical ranklets appear to perform better than tomographies and diagonal ranklets. The ROCs were obtained by varying the significance α of the single test. Results not included here clearly show that the ROC curves obtained using the same number of features drawn randomly from each category lead to inferior performances.

Combining horizontal and vertical ranklets (138 features or tests all together) leads to the highest classification accuracy (see Fig. 4; EER = 8%). In the same figure we also displayed the results obtained by using the first 72 principal components of the 138 features and measuring the Mahalanobis distance from the centroid. The PCA-based procedure appears to perform slightly better than the proposed technique (EER = 6%). This is consistent with the fact that the independence test

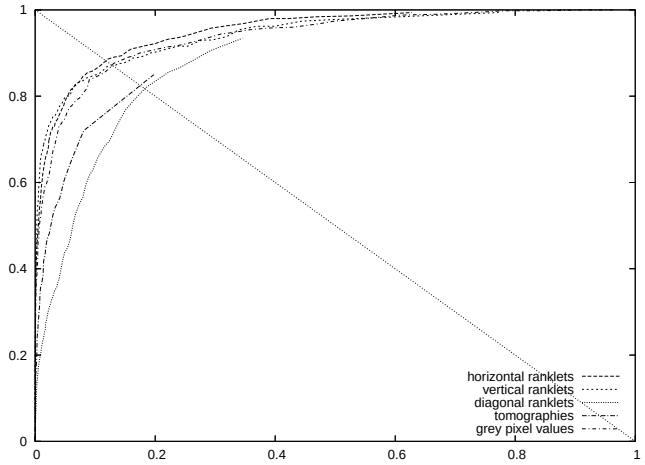

Figure 3: ROC curves on the MIT-CBCL test set. Comparison between grey values (where M has been set to $M = 75$), tomographies ($M = 38$), horizontal ($M = 56$), vertical ($M = 67$), and diagonal ($M = 70$) ranklets.

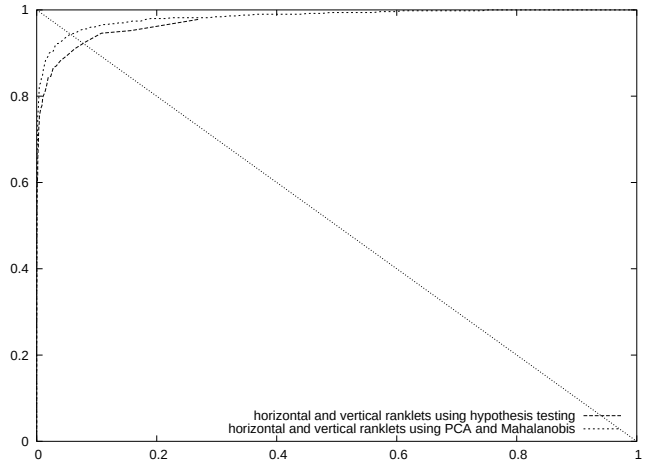

Figure 4: Hypothesis testing classification on horizontal and vertical ranklets combined ($M = 128$) compared with a PCA-based classifier over the same features (72 principal components, see text).

has been performed with a low threshold (to the purpose of retaining a not too small number of features).

6. Conclusions

We introduced a data–driven approach to feature selection and object detection based on hypothesis testing and positive examples only. We designed our hypothesis tests to be effective against the *omnibus* alternative by incorporating a form of the maximum likelihood principle into them. The saliency criterion for feature selection derives naturally from this design. The reported results support the potential of the proposed approach.

We believe that the main merit of this approach lies in the direct application of effective nonparametric statistical techniques with minimal assumptions on the probability distributions of the data. Clear strengths of this method are its generality, modularity, and wide applicability. On the other side, the flexibility of the approach can lead to suboptimal solutions unless some problem specific knowledge is injected into the system.

The experimental results we reported support the potential of our method; the EER of 8% is in line with the state of the art for this database [6, 2].

References

[1] Special issue on variable and feature selection. Journal on Machine Learning Reseach, march 2003.

[2] M. Alvira and R. Rifkin. An empirical comparison of SNoW and SVMs for face detection. Technical Report AI Memo 2001-004 - CBCL Memo 193, MIT, January 2001.

[3] E. L. Lehmann. *Nonparametrics: Statistical methods based on ranks*. Holden-Day, 1975.

[4] C. Papageorgiou and T. Poggio. A trainable system for object detection. *Internatonal Journal of Computer Vision*, 38(1):15–33, 2000.

[5] H. Schneiderman and T. Kanade. A statistical method for 3d object detection applied to faces and cars. In *Proc. IEEE Int Conf. CVPR*, 2000.

[6] F. Smeraldi. Ranklets: orientation selective nonparametric features applied to face detection. In *Proc. of the 16th ICPR, Quebec QC*, volume 3, pages 379–382, August 2002.

[7] S. Ullman, M. Vidal-Naquet, and E. Sali. Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, 5(7):67–79, 2002.

[8] P. Viola and M. Jones. Robust real-time object detection. In *II Int. Workshop on Stat. and Computat. Theories of vision - modeling, learning, computing and sampling*, 2001.

[9] A. Webb. *Statistical Pattern Recognition*. Oxford University Press, 1999.