Proceedings of the 16th International Conference on Pattern Recognition,
Quebec QC, vol. 3, pages 379–382, August 2002

Ranklets: orientation selective non-parametric features applied to face detection

Fabrizio Smeraldi
Halmstad University
Box 823, 30118 Halmstad, Sweden
fabrizio.smeraldi@ide.hh.se

Abstract

We introduce a family of multiscale, orientation-selective, non-parametric features ("ranklets") modelled on Haar wavelets. We clarify their relation to the Wilcoxon rank-sum test and the rank transform and provide an efficient scheme for computation based on the Mann-Whitney statistics. Finally, we show that ranklets outperform other rank features, Haar wavelets, SNoW and linear SVMs (based on independently published results) in face detection experiments over the $24'045$ *test images in the MIT-CBCL database.*

1. Introduction

The expression "non-parametrics" denotes statistical techniques that circumvent the problem of making assumptions about the underlying distribution of the data. To this category belong some of the classification algorithms recently developed within learning theory, such as Support Vector Machines (SVMs) [8]. However, the term is traditionally used in connection with statistical methods based on ranks [7].

Closely related to rank statistics are rank based features, that have been widely applied in the context of stereo correspondence [2, 6, 9] among others. Their main advantages consist in robustness to outliers and invariance under monotonic transformations, for example brightness and contrast changes and gamma correction.

In this paper we introduce a family of multiscale rank features ("ranklets") that show wavelet-style directional selectivity and are therefore well suited to characterise extended patterns with a complex geometry, such as for instance faces. We discuss the relation between our features, the rank transform of Zabih and Woodfill [9] and the Wilcoxon rank-sum test underlying it. We also provide an efficient scheme for the computation of ranklets based on the Mann-Whitney statistics.

We report face detection experiments over the $24'045$ test images of the MIT-CBCL face database. The performance of ranklets is compared with the rank and census transforms as well as with Haar wavelets, SVMs and the SNoW (Sparse Network of Winnows) algorithm, also according to results published by other research groups. Experiments show that ranklets significantly improve performance, thus proving to be a promising technique for pattern recognition on high noise, low resolution images.

2. The rank transform and other rank features

Given a set of N observations, by "ranking" we mean a permutation π of the integers from 1 to N that expresses the relative order of the observations. In this work we will be mainly concerned with a grey-level image I and we will indicate by $\pi^W(\vec{x})$ the rank of $I(\vec{x})$ among the intensity values of a suitably sized window W centred on pixel $\vec{x}$ (to simplify matters, we will assume that no two intensity values are equal; ties can be broken at random when they occur).

The rank transform [9] makes direct use of π by assigning to each pixel $\vec{x}$ the value of its rank: $\rho(\vec{x}) = \pi^W(\vec{x})$. This corresponds to the number of pixels in the local neighbourhood W whose intensity is lower than $I(\vec{x})$. The rank transform ρ is therefore a measure of the relative local brightness. A closely related similarity measure is Spearman's correlation coefficient r_s, that is proportional to the sum of the squared differences of rankings π_1^W, π_2^W over corresponding neighbourhoods of two images $I_1(\vec{x})$ and $I_2(\vec{y})$:

$$r_s \alpha \sum_i \left(\pi_1^W(\vec{x_i}) - \pi_2^W(\vec{y_i})\right)^2 \qquad (1)$$

Other rank features are defined directly in terms of pairwise comparisons of intensity values. We cite the census transform [9]: given a pixel $\vec{x}$ centred in W, let $\{\vec{x}_i\} = W \setminus \{\vec{x}\}$. Then the census transform associates to $\vec{x}$ the list $c_i(\vec{x})$ defined as follows: $c_i(\vec{x}) = 1$ if $I(\vec{x}) < I(\vec{x}_i)$, 0 otherwise. An example of a correlation measure based on pairwise comparisons is Kendall's τ [6].

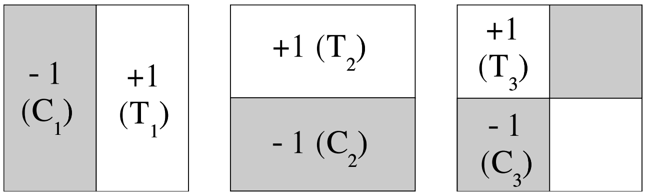

Figure 1. The three Haar wavelets $h_1(\vec{x})$, $h_2(\vec{x})$ and $h_3(\vec{x})$ (from left to right). Letters in parentheses refer to "treatment" and "control" pixel sets (see Section 4).

3. The Wilcoxon rank-sum test and the rank transform

The rank transform is closely related to the Wilcoxon rank-sum test for the comparison of two treatments [7]. Suppose that N quantities are split in two groups of n "treatment" and m "control" observations (according to the standard terminology). We are required to state whether the treatment observations are significantly higher than the controls. To this purpose we define the Wilcoxon rank sum statistics $\mathcal{W}_s$ as the sum of treatment ranks: $\mathcal{W}_s = \sum_{i=1}^{n} \pi(i)$. The treatment values are then judged to be significantly higher than the controls if the Wilcoxon statistics is above a critical value τ, $\mathcal{W}_s > \tau$. The value of τ determines the confidence level of the test.

The rank transform is equivalent to $\mathcal{W}_s$ when $n = 1$, that is when only one treatment observation is given. To show this, we identify the treatment observation with $I(\vec{x})$ and the controls with the $m = N - 1$ values $I(\vec{x}_i)$, where $\{\vec{x}_i\} \in W \setminus \{\vec{x}\}$ and W is an N-pixel window centred on $\vec{x}$. It follows that $\mathcal{W}_s = \pi^W(\vec{x}) = \rho(\vec{x})$. As a consequence, binarization of the rank transform by thresholding ($\rho(\vec{x}) \gtrless \tau$) has a specific statistical meaning, in that it amounts to fixing the critical value of the underlying Wilcoxon test. In other words, a pixel in the binarized rank transform is set to one if and only if its intensity is found to be larger than the intensity of the adjacent pixels in W with a confidence level specified by τ. Experimental results reported in Section 5 show that binarization of the rank transform can lead to improved performance.

4. Ranklets: a family of wavelet-style rank features

The close analogy between the Wilcoxon test and the rank transform can be carried further by devising new image descriptors that correspond to a number of "treatment" pixels n greater than 1. A convenient choice consists in splitting the N pixels in W in two groups of size $n = m = N/2$, thus assigning half of the pixels to the "treatment" group and half to the "control" group. This introduces a new degree of freedom, namely the geometric arrangement of the two regions in W. For any of the $\binom{N}{n}$ possible choices of treatment pixels, $\mathcal{W}_s$ will provide us with a different characterisation of the local neighbourhood. This wide arbitrariness can be exploited to obtain orientation selective features. To this purpose, we define the "treatment" and "control" groups starting from the three Haar wavelets [3] $h_j(\vec{x}), j = 1, 2, 3$ displayed in Figure 1. We identify the local neighbourhood W on which the ranking is performed with the support of the h_j. We then define the set of "treatment" pixels T_j as the counter-image of 1 under h_j: $T_j = h_j^{-1}(+1)$, and the set of "control" pixels C_j as the counter-image of -1: $C_j = h_j^{-1}(-1)$. For each partition of W, $W = T_j \cup C_j$, we then compute the value of the Wilcoxon statistics as $\mathcal{W}_s^j = \sum_{i=1}^{n} \pi^W(\vec{x}_i)$, with $\vec{x}_i \in T_j$.

We can conveniently replace $\mathcal{W}_s^j$ with the equivalent Mann-Whitney statistics $\mathcal{W}_{XY}^j = \mathcal{W}_s^j - n(n+1)/2$, which has an immediate interpretation in terms of pixel comparisons. As can be easily shown [7], $\mathcal{W}_{XY}^j$ is equal to the number of pixel pairs $(\vec{x}_p, \vec{y}_q)$ with $\vec{x}_p \in C_j$ and $\vec{y}_q \in T_j$ such that $I(\vec{x}_p) < I(\vec{y}_q)$. Its possible values therefore range from 0 to the number of pairs $(\vec{x}_p, \vec{y}_q) \in T_j \times C_j$, which is $mn = N^2/4$. Notice however that these pairwise comparisons are never carried out explicitly; the value of $\mathcal{W}_{XY}^j$ is obtained by ranking the pixels in W, which only requires $N \log N$ operations.

We can now define our image features, or "ranklets", as

$$\mathcal{R}_j = \frac{\mathcal{W}_{XY}^j}{mn/2} - 1. \tag{2}$$

The geometric interpretation of the $\mathcal{R}_j$ is straightforward in terms of the properties of $\mathcal{W}_{XY}^j$ and of the structure of the h_j. Consider for instance $\mathcal{R}_1$ and suppose that the local neighbourhood W straddles a vertical edge, with the darker side on the left (where C_1 is located) and the brighter side on the right (corresponding to T_1). Then $\mathcal{R}_1$ will be close to $+1$, as many pixels in T_1 will have higher intensity values than the pixels in C_1. Conversely, $\mathcal{R}_1$ will be close to -1 if the dark and bright side of the edge are reversed. Horizontal edges or other patterns with no global left-right variation of intensity will give a value close to zero. Therefore, $\mathcal{R}_1$ will respond to vertical edges in the images. By a similar argument $\mathcal{R}_2$ will detect horizontal edges, while $\mathcal{R}_3$ will be sensitive to corners formed by horizontal and vertical lines. These response patterns closely match those of the three Haar wavelets h_j.

Due to the close correspondence between Haar wavelets and ranklets, the multiscale nature of the former directly extends to the latter. To each translation and scaling of the h_j specified by $(\vec{x}_0, s)$ we associate the sets of treatment

Figure 2. Training faces (1st row), test faces (2nd row), and test non-faces (3rd row) from the MIT-CBCL set. Notice how a few of the test faces in the database are incorrectly framed (2nd row, right).

and control pixels defined by

$$T_{j;(\vec{x}_0,s)} = \{\vec{x} \mid h_j\left((\vec{x}-\vec{x}_0)/s\right) = +1\}\,, \tag{3}$$
$$C_{j;(\vec{x}_0,s)} = \{\vec{x} \mid h_j((\vec{x}-\vec{x}_0)/s) = -1\}\,. \tag{4}$$

We can now compute the value of $\mathcal{R}_j(\vec{x}_0, s)$ over the local neighbourhood $W_{(\vec{x}_0,s)} = T_j \cup C_j$, with $m = n = \#W_{(\vec{x}_0,s)}/2$.

5. Experimental results

We present the results of face detection experiments over the images of the MIT-CBCL face database [4], that consists of low-resolution grey level images ($19 \times 19 = 361$ pixels) of faces and non-faces. A training set of $2'429$ faces and a test set of 472 faces and $23'573$ non-faces are provided. All facial images nearly occupy the entire frame; considerable pose changes are represented (Figure 2). The negative test images were selected by a linear SVM classifier as those that looked most similar to faces among a much larger set of patterns [5]. The database also contains a set of $4'548$ non-faces intended for training. In our experiments we have discarded this training set of non-faces, since the notion of "non-face prototypes" appears to be problematic.

For the sake of simplicity, as well as to evidence the descriptive power of the features employed, we adopted a distance-based classification scheme. For each rank based or other transform, every image in the training and test sets is encoded as a (normalised) feature vector. A test image is recognised as a face if the distance from its corresponding vector to the closest face example is smaller than a threshold τ. The metric employed is the city block distance, that specialises to the Hamming distance for the case of the binarized rank transform.

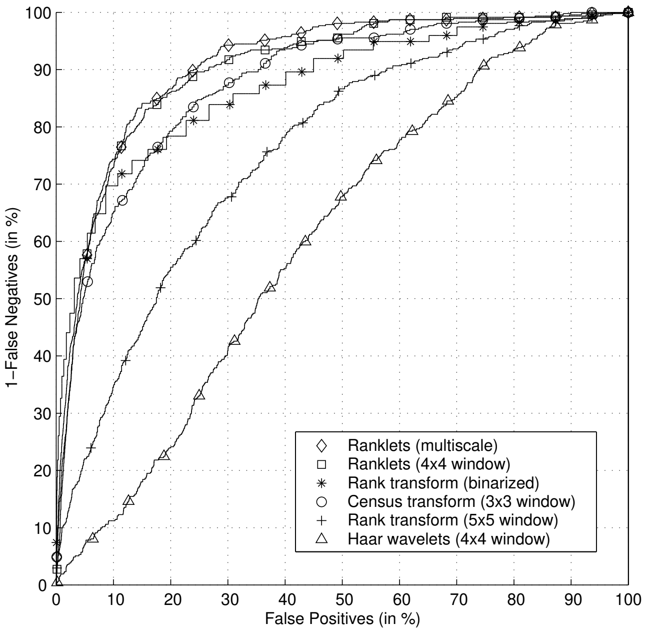

Figure 3. ROC curves for various types of features. The "step" appearance of the graph for the binarized rank transform is due to the limited range of values of the Hamming distance.

ROC curves for ranklets, the rank and census transforms and Haar wavelets are shown in Figure 3. Local neighbourhoods W of optimal size have been employed for ranklets and the rank transform; the choice of a 3×3 window for the census transform appeared to be natural. The window W has been centred at all image locations compatible with its size, yielding for instance 225 features for the rank transform and $256 \times 3 = 768$-dimensional feature vectors for the case of ranklets. In the case of multiscale ranklets, a total of $309 \times 3 = 927$ features has been extracted using 5 different sizes for W (note that we are mapping the originally 361-dimensional images to a higher dimensional feature space).

As can be seen, ranklets outperform all the other types of features we tested. The census transform and the binarized rank transform also yield good results. Remarkably, conventional Haar wavelets achieve a rather poor performance on these low-resolution, noisy images compared to the rank-based approaches. Equal Error Rates are reported in Table 1.

According to [1], both Linear Support Vector Machines and SNoW yield EERs in excess of 20% on the same database. Only polynomial SVMs show a performance comparable to ranklets, and this in spite of the fact that the $4'548$ negative training examples have been used (which is not true of our case).

Type of features	EER
Ranklets (multiscale)	15.9 %
Ranklets (4x4 window)	16.7 %
Census transform	20.3 %
Rank transform (binarized)	21.6 %
Rank transform	31.4 %
Haar wavelets	41.8 %

Table 1. EER as a function of the features employed.

6. Conclusions

We have introduced a new family of rank features, called "ranklets". Closely modelled on Haar wavelets, ranklets inherit from them the orientation selectivity and the multiscale nature. Their definition in terms of the Mann-Whitney statistics provides a connection to the Wilcoxon rank-sum test, an efficient computing scheme and an intuitive interpretation in term of pairwise comparisons of pixel intensity values.

Experimental results over a test set of $24'045$ images show that ranklets outperform a wide range of other algorithms, notably including Haar wavelets, SNoW and linear SVMs applied directly to the intensity data. In future work we plan to investigate the behaviour of SVMs applied to the classification of feature vectors of ranklets.

Acknowledgements

Thanks to E. Franceschi, F. Isgrò and A. Verri for discussion on the topics of this paper.

References

[1] 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. http://www.ai.mit.edu/projects/cbcl.

[2] D. N. Bhat and S. K. Nayar. Ordinal measures for visual correspondence. In *Proceedings of CVPR*, pages 351–357, 1996.

[3] I. Daubechies. *Ten lectures in wavelets*. Society for industrial and applied mathematics, Philadelphia, USA, 1992.

[4] M. C. for Biological and C. Learning. CBCL face database no. 1. http://www.ai.mit.edu/projects/cbcl.

[5] B. Heisele, T. Serre, S. Mukherjee, and T. Poggio. Feature reduction and hierarchy of classifiers for fast object detection in video images. In *Proceedings of CVPR'01, Kauai (Hawaii)*, volume 2, pages 18–24, December 2001.

[6] M. Kendall and J. D. Gibbons. *Rank correlation methods*. Edward Arnold, 1990.

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

[8] V. N. Vapnik. *The nature of statistical learning theory*. Springer–Verlag, 1995.

[9] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In *Proceedings of the 3rd ECCV*, pages 151–158, 1994.