Face Recognition Using Ordinal Features[*]

ShengCai Liao, Zhen Lei, XiangXin Zhu, ZheNan Sun, Stan Z. Li, and Tieniu Tan

Center for Biometrics and Security Research & National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences,
95 Zhongguancun Donglu Beijing 100080, China
`http://www.cbsr.ia.ac.cn`

Abstract. In this paper, we present an ordinal feature based method for face recognition. Ordinal features are used to represent faces. Hamming distance of many local sub-windows is computed to evaluate differences of two ordinal faces. AdaBoost learning is finally applied to select most effective hamming distance based weak classifiers and build a powerful classifier. Experiments demonstrate good results for face recognition on the FERET database, and the power of learning ordinal features for face recognition.

1 Introduction

It is believed that the human vision system uses a series of levels of representation, with increasing complexity. A recent study on local appearance or fragment (or local region) based face recognition [7] shows that features of intermediate complexity are optimal for basic visual task of classification, and mutual information for classification is maximized in a middle range of fragment size. Existing approaches suggest a trade-off between the complexity of features and the complexity of the classification scheme. Using fragment features is advantageous [8] in that they reduce the number of features used for classification for richer information content of the individual features, and that a linear classifier may suffice when proper fragment features are selected; on the other hand, with simple generic features, the classifier has to use higher-order properties of their distributions.

However, whether to use fragment or generic features remain a question. While using fragment features may be advantages for classification between apparently different classes, such as between a car and a face, the conclusion may not apply for object classes in which the differences in their appearances are not so obvious, eg faces of different individuals. For the latter case, more elementary and generic feature should provide better discriminative power. This in general requires a nonlinear classifier in which higher order constraints are incorporated.

In this regard, we consider a class of simple features: the ordinal relationship. Ordinal features are defined based on the qualitative relationship between two image regions and are robust against various intra-class variations [3, 5, 6]. For example, they invariant to monotonic transformations on images and is flexible enough to represent different local

---

structures of different complexity. Sinha [5] shows that several ordinal measures on facial images, such as those between eye and forehead and between mouth and cheek, are invariant with different persons and imaging conditions, and thereby develops a ratio-template for face detection. Schneiderman [4] uses an ordinal representation for face detection.

Face recognition is a more difficult problem than face detection. While ordinal features have shown excellent separability between the face class and the rest of the world, it remains a question whether it is powerful enough for face recognition [6]. Thoresz [6] believes ordinal features are only suited for face detection but too weak for fine discrimination tasks, such as personal identification.

In this paper, we present an ordinal feature based method for face recognition. Ordinal features are generated using ordinal filters and are used to represent faces. Hamming distance of many local sub-windows is computed to evaluate differences of two ordinal faces. AdaBoost learning is finally applied to select most effective ordinal features and build a powerful classifier. Experiments demonstrate good results for face recognition on the FERET database.

The contributions of this work are summarized as follows: While ordinal features have been used for face detection, its application in face recognition is for the first time. We will show that ordinal features when properly selected using statistical learning method can do well for face based personal identification. The second contribution is that unlike manual feature selection as in [5], we propose to use a statistical learning method for selecting effective ordinal features and thereby constructing a strong classifier for face recognition.

The rest of this paper is organized as follows. In Section 2, we introduce ordinal features. In Section 3, AdaBoost learning is applied to select most discriminative feature, while removing large redundance in the feature set, and learn boosted classifiers. Section 4 describes weak classifiers for ordinal features learning. Experimental results are presented in Section 5.

2   Ordinal Features

Ordinal features come from a simple and straightforward concept that we often use. For example, we could easily rank or order the heights or weights of two persons, but it is hard to answer their precise differences. For computer vision, the absolute intensity information associated with an face can vary because it can changes under various illumination settings. However, ordinal relationships among neighborhood image pixels or regions present some stability with such changes and reflect the intrinsic natures of the face.

An ordinal feature encodes an ordinal relationship between two concept. Figure 1 gives an example in which the average intensities between regions A and B are compared to give the ordinal code of 1 or 0. Ordinal features are efficient to compute. Moreover, the information entropy of the measure is maximized because the ordinal code has nearly equal probability of being 1 or 0 for arbitrary patterns.

While differential filters, such as Gabor filters, are sufficient for comparison of neighboring regions', Balas and Sinha [1] extend those filters to "dissociated dipoles" for

Fig. 1. Ordinal measure of relationship between two regions. An arrow points from the darker region to the brighter one. Left: Region A is darker than B, i.e. $A \prec B$. Right: Region A is brighter than B, i.e. $A \succ B$.
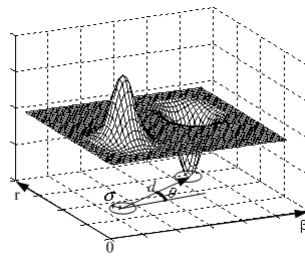

Fig. 2. Dissociated dipole operator

non-local comparison, shown in Figure 2. Like differential filters, a dissociated dipole also consists an excitatory and an inhibitory lobe, but the limitation on the relative position between the two lobes is removed. There are three parameters in dissociated dipoles:

– The scale parameter σ: For dipoles with a Gaussian filter, the standard deviation σ is an indicator of the scale.
– The inter-lobe distance d: This is defined as the distance between the centers of the two lobes.
– The orientation θ: This is the angle between the line joining the centers of the two lobes and the horizontal line. It is in the range from 0 to 2π.

We extend dissociated dipoles to dissociated multi-poles, as shown Figure 3. While a dipole tells us the orientation of a slope edge, a multi-pole can represent more complex image micro-structures. A multi-pole filter can be designed for a specific macro-structure, by using appropriate lobe shape configuration. This gives much flexibility for filter design.

To be effective for face recognition or image representation, there are three rules in development of dissociated multi-poles (DMPs):

– Each lobe of a DMP should be a low-pass filter. On one hand, the intensity information within the region of the lobe should be statistically estimated; on the other hand, the image noise should attenuated by low-pass filtering.
– To obtain the locality of the operator, the coefficients of each lobe should be arranged in such a way that the weight of a pixel is inverse proportional to its distance from the lobe center. Gaussian mask satisfies this; there are other choices as well.
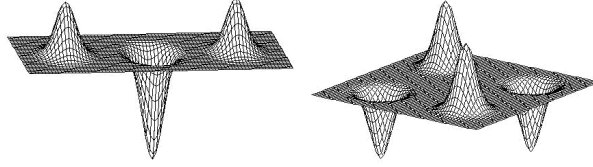
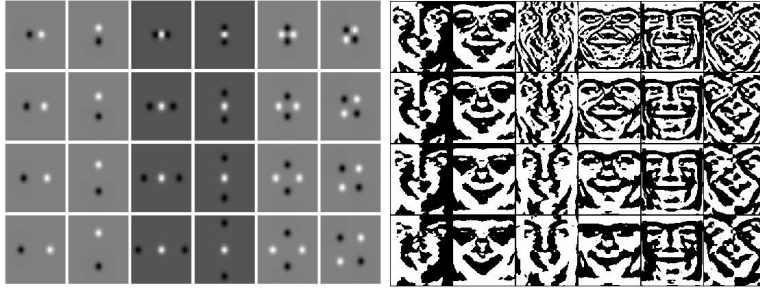**Fig. 3.** Dissociated multi-pole: tri- and quad-pole filters

Fig. 4. The 24 ordinal filters used in the experiments, and the corresponding filtered images of a face

- The sum of all lobes' coefficients should be zero, so that the ordinal code of a non-local comparison has equal probability being 1 or 0. Thus the entropy of a single ordinal code is maximized. In the examples shown in Figure 3, the sum of two excitatory lobes' weights is equal to the inhibitory lobes' total absolute weights.

In this paper, we use 24 disassociated multi-pole ordinal filters as shown in Fig.4. The filter sizes are all 41x41 pixels. The Gaussian parameter is uniformly $\sigma = \pi/2$. The inter-pole distances are $d = 8, 12, 16, 20$ for the 2-poles and 4-poles, and $d = 4, 8, 12, 16$ for the 3-poles. For 2-poles and 3-poles, the directions are 0 and $\pi/2$; for the 4-poles, the directions are 0 and $\pi/4$. A more complete set would include a much larger number of filters with varying parameters. Optimization of parameters would take into consideration of the final performance as well as costs in memory and training speed.

3   AdaBoost Learning

Because the large Ordinal feature set contains much redundant information, A further processing is need to remove the redundancy and build effective classifiers. This is done in this work by using the following AdaBoost algorithm [2]:

Input: Sequence of N weighted examples

$$\{(x_1, y_1, w_1), (x_2, y_2, w_2), \ldots, (x_n, y_n, w_n)\};$$

Initial distribution P over the n examples;

Weak learning algorithm **WeakLearn**;

Integer T specifying number of iterations;

Initialize $w_i^1 = P(i)$ for $i = 1, \ldots, n$;

For $t = 1, \ldots, T$:

    1. Set $p_i^t = w_i^t / \sum_i w_i^t$;

    2. Call **WeakLearn**, providing it with the distribution p;

       get back hypothesis $h_t(x_i) \in \{0, 1\}$ for each x_i;

    3. Calculate the error of h_t : $\epsilon_t = \sum_{i=1}^{N} p_i^t |h_t(x_i) - y_i|$;

    4. Set $\beta_t = \frac{\epsilon_i}{(1-\epsilon_t)}$;

    5. Set the new weights to $w_i^{t+1} = \beta_i^{1-|h_t(x_i)-y_i|}$;

Output the hypothesis

$$H(x) = \begin{cases} 1 \text{ if } \sum_{t=1}^{T} \left(\log \frac{1}{\beta_t} \right) h_t(x) \geq \sum_{t=1}^{T} \left(\log \frac{1}{\beta_t} \right) \\ 0 \text{ otherwise} \end{cases}$$

AdaBoost iteratively learns a sequence of weak hypotheses $h_t(x)$ and linearly combines them with the corresponding learned weights $\log \frac{1}{\beta_t}$. Given a data distribution p, AdaBoost assumes that a **WeakLearn** procedure is available for learning a sequence of most effective weak classifiers $h_t(x)$. This will be discussed in the next section.

4   Weak classifiers

The simplest weak classifier can be constructed for each pixel and each filter type, which we call single bit weak classifier (SBWC). We can concatenate all the filtered images into a complete filtered image. Consider every pixel in the complete image as a bit. An SBWC outputs 0 or 1 according to the bit value. At each iteration, the AdaBoost learning selects the bit by which the performance is the best, eg causing the lowest weighted error over the training set.

A more involved weak classifier can be designed based on a spatially local subwindow instead of a single bit. The advantage is that some statistic over a local subwindow can be more stable than that at a bit. In this scheme, the Hamming distance can be calculated between the ordinal values in the two corresponding subwindows. The Hamming distance as a weak classifier can be used to make a weak decision for the classification. The use of subwindows gives one more dimension of freedom, the subwindow size. A different size leads to a different weak classifier. The two types of week classifiers will be evaluated in the experiment.

5   Experiments

The proposed method is tested on the FERET face database. The training set contains 540 images from 270 subjects. The test set contains 1196 galleries and 1195 probes from 1196 subjects. All images are cropped to 142 pixels high by 120 pixels wide, according to the eyes positions. The 24 ordinal filters are applied to all the images.
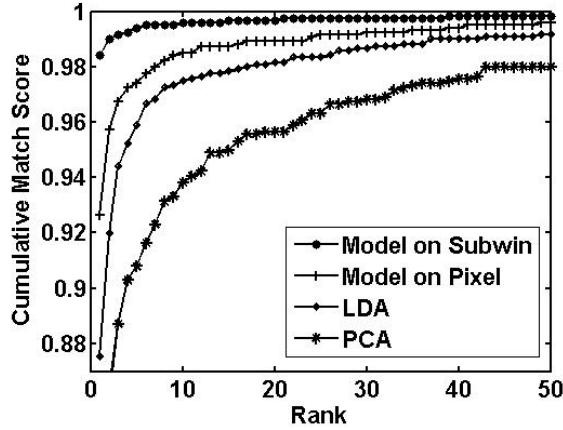
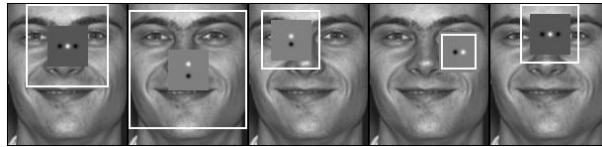**Fig. 5.** Cumulative match curves of 4 compared methods

Fig. 6. The first 5 features and associated subwindow sizes selected by AdaBoost learning

The experiments evaluate the two AdaBoost learning based methods. The first is with the SBWC for feature selection and classifier learning. The second uses the local subwindow of ordinal features to construct Hamming distance based weak classifiers for AdaBoost learning. These two methods are be compared with the standard PCA and LDA methods (derived using the intensity images).

For the first method, a total of 173 weak classifiers, are trained to reach the training error rate of zero on the training set.

For the second method, 20 subwindow sizes are used: 6x6, 12x12, ..., 120x120 where the length of the side is incremented by 6. A single strong classifiers, consisting of 34 weak classifiers, is trained to reach the error rate of zero on the training set. The first 5 learned weak classifiers are shown in Fig.6. In the figure, the type of the filter and the subwindow size indicates the corresponding weak classifier.

Figs.5 describes the performances of the tested methods, in terms of the accumulative match curves, where the first method is named "Model on Pixel" and the second named "Model on Subwin". "Model on Subwin" performs the best, "Model on Pixel" the second, followed by LDA and PCA. The rank one recognition rates for the four methods are 98.4%, 92.5%, 87.5%, and 80.0%, respectively. This shows that the methods based ordinal features with statistical learning give good face recognition performances. Of the two proposed methods, "Model on Subwin" method is obviously advantageous: it needs fewer weak classifiers yet achieves a very good result.

6   Summary and Conclusions

In this paper, we have proposed a learning method for ordinal feature based face recognition. While it was believed that ordinal features were only suited for face detection and too weak for fine discrimination tasks, such as personal identification that[6], our preliminary results show that ordinal features with statistical learning can be powerful enough for complex tasks such as personal identification. In the future, we will investigate the effects of varying ordinal filter parameters, and find how intermediate features such as fragments can be built based on the simple ordinal features, and how to construct higher order ordinal features effectively using statistical learning.

References

1. B. Balas and P. Sinha. "Toward dissociated dipoles: Image representation via non-local comparisons". CBCL Paper #229/AI Memo #2003-018, MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, USA, August 2003.
2. Y. Freund and R. Schapire. "A decision-theoretic generalization of on-line learning and an application to boosting". *Journal of Computer and System Sciences*, 55(1):119–139, August 1997.
3. J. Sadr, S. Mukherjee, K. Thoresz, , and P. Sinha. "Toward the fidelity of local ordinal encoding". In *Proceedings of the Fifteenth Annual Conference on Neural Information Processing Systems*, Vancouver, British Columbia, Canada, December 3-8 2001.
4. H. Schneiderman. "Toward feature-centric evaluation for efficient cascaded object detection". In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1007–1013, Washington, DC, USA, June 27 - July 2 2004.
5. P. Sinha. "Toward qualitative representations for recognition". In *Proceedings of the Second International Workshop on Biologically Motivated Computer Vision*, pages 249–262, Tubingen, Germany, November 22-24 2002.
6. K. J. Thoresz. On qualitative representations for recognition. Master's thesis, MIT, July 2002.
7. S. Ullman, M. Vidal-Naquet, and E. Sali. "Visual features of intermediate complexity and their use in classification". *Nature Neuroscience*, 5(7), 2002.
8. M. Vidal-Naquet and S. Ullman. "Object recognition with informative features and linear classification". In *Proceedings of IEEE International Conference on Computer Vision*, Nice, France, 2003.