Structured Ordinal Features for Appearance-Based Object Representation

Shengcai Liao, Zhen Lei, Stan Z. Li, Xiaotong Yuan,
and Ran He

Center for Biometrics and Security Research & National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences,
95 Zhongguancun Donglu, Beijing 100080, China
{scliao,zlei,szli,xtyuan,rhe}@nlpr.ia.ac.cn
http://www.cbsr.ia.ac.cn

Abstract. In this paper, we propose a novel appearance-based representation, called Structured Ordinal Feature (SOF). SOF is a binary string encoded by combining eight ordinal blocks in a circle symmetrically. SOF is invariant to linear transformations on images and is flexible enough to represent different local structures of different complexity. We further extend SOF to Multi-scale Structured Ordinal Feature (MSOF) by concatenating binary strings of multi-scale SOFs at a fix position. In this way, MSOF encodes not only microstructure but also macrostructure of image patterns, thus provides a more powerful image representation. We also present an efficient algorithm for computing MSOF using integral images. Based on MSOF, statistical analysis and learning are performed to select most effective features and construct classifiers. The proposed method is evaluated with face recognition experiments, in which we achieve a high rank-1 recognition rate of 98.24% on FERET database.

1 Introduction

Object recognition from images is a challenging problem in computer vision. The main difficulties arise due to many uncertainties such as viewpoint and illumination changes. To overcome such problems, appearance-based object representation has been a hot issue in the past two decades. Among these, PCA [17] and LDA [2] are two classical linear methods that have significantly advanced object recognition techniques. But linear, holistic appearance-based methods can not capture subtleties of various objects, and holistic features are unstable under various illumination changes. It is believed that localized appearance-based features, which reflect the intrinsic properties of an object, can be more powerful for object recognition. Thus local features have been investigated a lot by researchers in recent years, such as Local feature analysis (LFA) [10], Gabor wavelet-based features [5,19,7], Local Binary Patterns (LBP) [1], and ordinal measures [15].

Local Binary Pattern (LBP) is a powerful local descriptor for microfeatures of images [9]. The LBP operator labels the pixels of an image by thresholding the 3×3-neighborhood of each pixel with the center value and considering the result as a binary number. Ahonen *et al.* proposed a novel approach for face recognition, which takes advantage of the Local Binary Pattern (LBP) histogram [1]. However, the original LBP

S.K. Zhou et al. (Eds.): AMFG 2007, LNCS 4778, pp. 183–192, 2007.

has its small spatial support area, hence the bit-wise comparison therein made between two single pixel values is much affected by noise. Moreover, features calculated in the local 3×3 neighborhood cannot capture larger scale structure (macrostructure) that may be dominant features of objects.

Recently, ordinal measure is discussed frequently as a method for representing local image structures. Ordinal features are defined based on the qualitative relationship between two image regions and are robust against various intra-class variations [11,15,16]. For example, they are invariant to linear transformations on images and is flexible enough to represent different local structures of different complexity. Sinha [15] 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 [13] also uses an ordinal representation for face detection. While in the task of face recognition, which is a more complex problem than face detection, Thoresz [16] believes that ordinal features are not suited because they are too weak. Yet Liao *et al.* propose an ordinal feature based face recognition method for the first try, and obtained a promising results [6].

In this work, we propose a novel representation, called Structured Ordinal Feature (SOF). It is believed that the human vision system uses a series of levels of representation, with increasing complexity. Since one single ordinal feature is too simple to represent complex structures, we propose to combine several ordinal measures together to form a more powerful encoding of local image structures. SOF is a binary string encoded by combining eight ordinal blocks in a circle symmetrically. Using integral images, the comparison of average intensities between two blocks can be calculated very efficiently. Furthermore, Multi-scale Structured Ordinal Feature (MSOF) can be derived via concatenating binary strings of multi-scale SOFs at a fix position. This way, MSOF encodes not only microstructure but also macrostructure of image patterns, thus provides a more complete image representation. Based on MSOF, we define several dissimilarity measures, and perform statistical learning to select the most effective features and construct classifiers. Finally, we apply it to face recognition to illustrate the power and effectiveness of our proposed method.

The rest of this paper is organized as follows: In Section 2, the SOF and MSOF representations are introduced, and several dissimilarity measures based on MSOF are defined for discrimination tasks. In Section 3, statistical learning is applied for MSOF-based feature selection and classifier construction. Later, experiments with face recognition are shown in Section 4, and finally we conclude this paper in Section 5.

2 Structured Ordinal Feature

2.1 Ordinal Feature

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 object can vary because it can change 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 object.

An ordinal feature encodes an ordinal relationship between two concept. Fig.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. The information entropy of the ordinal measure is maximized because the ordinal code has nearly equal probability of being 1 or 0 for arbitrary patterns.

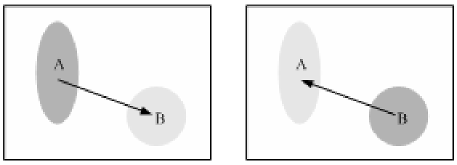

Fig. 1. Ordinal measure of relationship between two regions. An arrow points from the darker region to the brighter one. Left: Region A is darker than B, *i.e.* $A \prec B$. Right: Region A is brighter than B, *i.e.* $A \succ B$.

2.2 Structured Ordinal Feature

It is believed that the human vision system uses a series of levels of representations, with increasing complexity. Since one single ordinal feature is too simple to represent complex structures, we propose to combine several ordinal measures together to form a more powerful encoding of local image structures. We call this combination of ordinal encoding Structured Ordinal Feature (SOF). SOF is a binary string encoded by combining eight square ordinal blocks in a circle symmetrically, which is inspired by the encoding of local binary patterns [9]. Fig.2 illustrates an example of how SOF encoded.

There are two parameters in SOF: one is the size s of the square blocks, the other is the radius r of the circle. The parameter pair (s, r) denotes the scale of SOF. An SOF feature of scale (s, r) at pixel location (x, y) can be denoted as $SOF_{s,r}(x, y)$.

Structured Ordinal Feature extends the original ordinal feature, and can be used to represent various image structures, which may be some intrinsic properties of image object. The original ordinal feature can only show the contrast information between two regions, while using SOF, more local image structures can be represented. Fig. 3 shows some image structures that can only be represented by SOF. These structures are basic properties within many image objects. Therefore, SOF provides a more efficient and flexible way for appearance-based object representation.

Note that the scalar values of averages over blocks can be computed very efficiently [14] from the summed-area table [3] or integral image [18]. For this reason, the computation of SOF is very fast. Also note that the comparison result between averages of two blocks is invariant when the image is linearly transformed, thus the encoding of SOF is invariant to linear transformations on images.

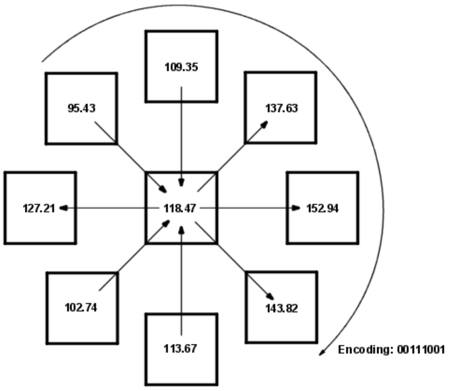

Fig. 2. An example of SOF encoding. Eight square ordinal blocks are combined in a circle symmetrically to form a structured filter. The number in each block is the average intensity within the corresponding image region. The arrows represent the ordinal relationships. According to these relationships, the result is encoded as a binary string.

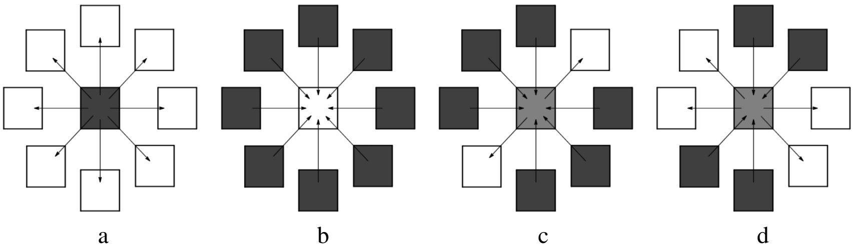

Fig. 3. Some image structures represented by SOF. a. Centered darker region; b. Centered brighter region; c. Brighter strip; d. Cross strips.

2.3 Multi-scale Structured Ordinal Feature

To construct a more complete image representation, we develop the operator of multi-scale SOF (MSOF). MSOF can be derived via concatenating binary strings of multi-scale SOFs at a fix position. See Fig.4 for an example. Suppose there are n scales of SOF, then an MSOF at location (x, y) is encoded as

$$MSOF(x,y) = SOF_{s_0,r_0}(x,y) \oplus SOF_{s_1,r_1}(x,y) \oplus \cdots \oplus SOF_{s_{n-1},r_{n-1}}(x,y),$$

where $\oplus$ denotes binary concatenation operator. In this way, MSOF encodes not only microstructure but also macrostructure of image patterns, thus provides a more complete image representation.

The scale parameters of MSOF should be carefully designed so that the operator will cover the neighborhood as well as possible while minimizing the amount of redundant information. Consequently, the square blocks are set to fill the eight directions well within the same circle, while touching each other as well as possible between circles.

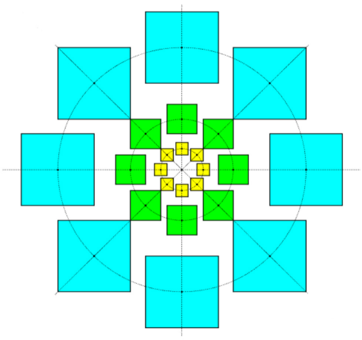

Fig. 4. An example of MSOF operator. Three scales of SOFs are combined at the center. The encodings are concatenated to form a binary string of 24 bits.

2.4 Dissimilarity Measure

For discrimination tasks, the Hamming distance can be used between two MSOFs at the same position. Since the encoding of MSOF is a concatenated binary string at a fix position, Hamming distance between two MSOFs at the same position can be used to measure the difference of two images at that position. Let $Hamm(\cdot,\cdot)$ denote the Hamming distance between two binary strings, then distance between two MSOFs at location (x,y) of image I' and I'' can be measured as

$$d_{hamm}(x,y) = Hamm(MSOF'(x,y), MSOF''(x,y)) \tag{1}$$

This kind of difference measures the percentage of bitwise difference between two MSOFs, ranging from 0 to 1.

Consequently, Hamming distances at all positions provides a discriminative feature set for object recognition, and classifiers can be further developed based on these features. One simple classifier is Nearest Neighbor (NN) classifier, using MSOF distance summed over the whole image as a dissimilarity measure. Suppose all images are H pixels high and W pixels wide, then such dissimilarity measure between two images I' and I'' can be defined as

$$d_{hamm} = \frac{1}{H \times W} \sum_{x,y} d_{hamm}(x,y) \tag{2}$$

The NN classifier can be used to provide a baseline performance. The latter section will demonstrate promising results given by MSOF-based NN classifier.

Another discriminative feature set can be constructed considering spacial information. Let $B = \{B_0, B_1, \cdots, B_{m-1}\}$ be a set of m-1 blocks of various sizes and positions over the whole image, then a set of features containing local region information can be defined as

$$d_{hamm}(B_i) = \frac{1}{H_i \times W_i} \sum_{(x,y)\in B_i} d_{hamm}(x,y), \quad i = 0, 1, \cdots, m-1, \tag{3}$$

where H_i and W_i are height and width of block B_i respectively. Accordingly m discriminative features are defined for object recognition. We call them block-based MSOF Hamming dissimilarity measures.

Hamming distance based MSOF dissimilarity measure is effective when images are well aligned. However, based on pixel-wise comparison, Hamming distances are influenced by image misalignment. For more robust and efficient discrimination, we construct a feature set based on spatially distributed histograms, which are also used in [1]. Using the above notations, the MSOF histogram is defined as

$$\begin{aligned} Hist(B_i, j) = \frac{1}{H_i \times W_i \times n} \sum_{(x,y) \in B_i} \sum_{k=0}^{n} I\{SOF_{s_k, r_k}(x, y) = j\}, \\ i = 0, 1, \cdots, m-1,\ j = 0, 1, \cdots, L-1, \end{aligned} \tag{4}$$

where j is an SOF code, and $L = 2^8$, thus the histogram has 256 bins. Based on these spatially distributed histograms, discriminative features can be defined as difference between two corresponding histogram bins:

$$\begin{aligned} d_{hist}(B_i, j) = |Hist'(B_i, j) - Hist''(B_i, j)|, \\ i = 0, 1, \cdots, m-1, j = 0, 1, \cdots, L-1. \end{aligned} \tag{5}$$

We call the above dissimilarity measure histogram-based MSOF dissimilarity measure. These features provide local histogram information and hence will be more efficient for discrimination tasks. However, considering blocks of all sizes and locations, the feature dimensions will be very large ($m \times 256$). Therefore, a proper technique should be used to reduce the dimension and construct classifiers.

3 Statistical Learning for Object Recognition

The above MSOF-based dissimilarity measures provide an over-complete discriminative feature set. The only question remained is how to use them to construct a powerful classifier. Because those excessive measures contain much redundant information, a further processing is needed to remove the redundancy and build effective classifiers. In this paper we use Gentle AdaBoost algorithm [4] to select the most effective MSOF-based dissimilarity measures.

Boosting can be viewed as a stage-wise approximation to an additive logistic regression model using Bernoulli log-likelihood as a criterion [4]. Developed by Friedman et al, Gentle AdaBoost modifies the popular version of the Real AdaBoost procedure [12], using Newton stepping rather than exact optimization at each step. Empirical evidence suggests that Gentle AdaBoost is a more conservative algorithm that has similar performance to both the Real AdaBoost and LogitBoost algorithms, and often outperforms them both, especially when stability is an issue.

While an AdaBoost procedure essentially learns a two-class classifier, we convert the multi-class problem into a two-class one using the idea of intra- and extra-class difference [8]. However, here the difference data are derived from the MSOF-based dissimilarity measures rather than from the images. An MSOF-based dissimilarity measure is taken between two MSOF representations, which is intra-class if the two images are of the same class, or extra-class if not. The MSOF-based dissimilarity measures are used to construct weak classifiers for the above AdaBoost learning. The best current weak classifier is the one for which the weighted intra-class MSOF-based dissimilarity measure (over the training set) is minimized while that of the extra-class is maximized. After AdaBoost learning, the feature dimensions of the MSOF-based dissimilarity measures are dramatically reduced, meanwhile a powerful classifier is constructed.

With the two-class scheme and the learned classifier, the object recognition procedure will work in the following way: It takes a probe image and a gallery image as the input, and computes a feature vector from the two images using the selected MSOF-based dissimilarity measures, then it calculates a similarity score for the feature vector using the learned AdaBoost classifier. Finally a decision is made based on the score, to classify the feature vector into the intra-class (the same objects) or the extra-class (different objects).

4 Experiments

To evaluate the proposed MSOF-based representation, we apply it to face recognition, which is a typical object discrimination task and also a hard problem. Experiment is evaluated on FERET fa/fb face database. Face images are cropped into 150 pixels high and 130 pixels wide, according to their eye coordinates. The non-face area is excluded using an elliptical mask, and the gray histogram within the elliptical mask is equalized. The FERET training CD contains 1002 frontal face images from 429 subjects. The test set contains 1196 galleries and 1195 probes from 1196 subjects.

First, we want to show some basic effects of MSOF filtering. Fig. 5 shows three examples of face images, and Fig. 6 demonstrates the corresponding filtered images, each of which is filtered by a 5-scale MSOF operator. The scale parameters we use in this paper are: (1,1), (3,3), (5,7), (9,14), (17,27), which are determined according to the principle we mentioned in Section 2.3. From Fig. 6 we could see that large scale of SOF encodes macrostructure of image objects, while small scale reveals fine details of local structures. Hence combining all these scales, MSOF provides a complete object representation.

Dissimilarity measures of a pair of intra-personal images and a pair of extra-personal ones are shown in Fig. 5. The difference image is generated using Equ. (1), where the brighter pixels indicate larger differences. From Fig. 5 we could see clearly that images of extra-personal pair have larger differences than that of intra-personal pair. Using Equation of (2), we could exactly measure the dissimilarity between two images. In this example, the intra-personal dissimilarity is 0.1603, while the extra-personal one is 0.2974. It follows that the MSOF-based dissimilarity measure is able to provide promising power for discriminating intra-/extra-class differences.

The next experiment is designed to evaluate the baseline performance of NN classifier using dissimilarity measure of Equ. (2). We follow the standard FERET test protocol of fa/fb face database, which contains 1196 gallery images and 1195 probe images. One advantage of MSOF-based NN classifier is that it could be directly used without training. The cumulative match score curve is shown in Fig. 7. The rank-1 recognition rate is 80%, a promising result.

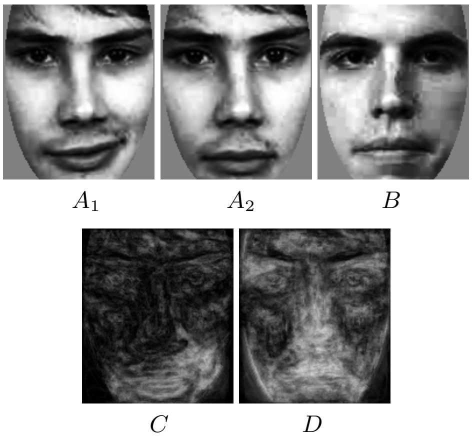

Fig. 5. Examples of dissimilarity measures based on MSOF. The first row: original face images, where A_1 and A_2 are the same person, and B is another person. The second row: difference images calculated via Equ. (1), where C is generated using A_1 and A_2, and D is generated using A_2 and B. The total dissimilarity of (A_1, A_2) pair computed with Equ. (2) is 0.1603, while that of (A_2, B) pair is 0.2974.

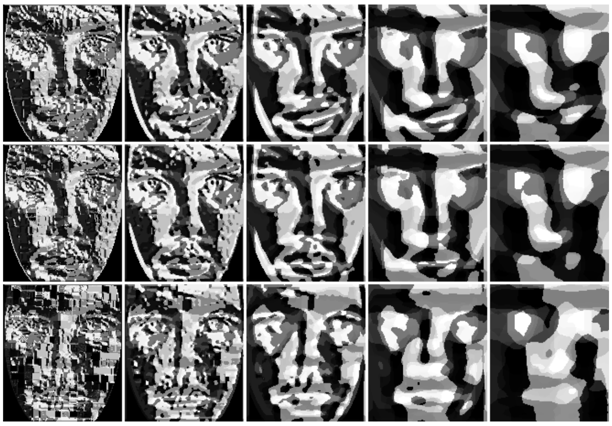

Fig. 6. Examples of MSOF filtered images with 5 scales. Each column is corresponding to one scale, which becomes larger from left to right. For each scale, 8-bits binary string of each SOF encoding is converted to a decimal number ranging from 0 to 255, which is displayed as a pixel label here. Each row is generated with one face image. The three rows are corresponding to image A_1, A_2, and B in Fig. 5 respectively.

Finally, we train two AdaBoost classifiers based on MSOF dissimilarity measure, using the FERET training set of 1002 images. The first AdaBoost classifier is constructed using block-based MSOF Hamming dissimilarity measure (Equ. (3)), and the second one is using histogram-based MSOF dissimilarity measure (Equ. (5)). The results are

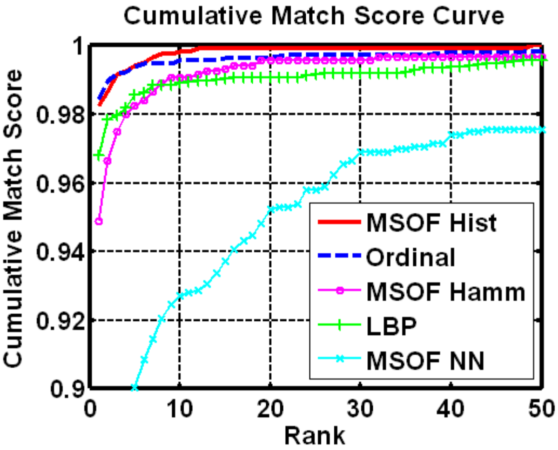

Fig. 7. Cumulative match score curves. "MSOF Hist" is for histogram-based MSOF dissimilarity measure (Equ. (5)); "MSOF Hamm" is for block-based MSOF Hamming dissimilarity measure (Equ. (3)); "MSOF NN" is for Nearest Neighbor classifier using dissimilarity measure of Equ. (2); "Ordinal Feature" is the result of [6], and "LBP" is that of [1].

compared with the approach based on LBP [1] and Ordinal Feature [6] shown in Fig. 7. From the cumulative match score curve, we could see that classifier using histogram-based MSOF dissimilarity measure outperforms all other algorithms, though Ordinal Feature of [6] is slightly better before rank3. The rank-1 recognition rate of histogram-based MSOF is 98.24%, which is an excellent result on FERET database. The performance of block-based MSOF Hamming dissimilarity measure classifier is not the best, but it is also outperforms LBP after rank-8, and it achieves high recognition rates near that of Ordinal Feature after rank-20.

5 Summary and Conclusions

This paper proposes a novel appearance-based representation, called Structured Ordinal Feature (SOF). We show that SOF is invariant to linear transformations on images and can be efficiently computed. It can be further extended to Multi-scale Structured Ordinal Feature (MSOF) to encode both microstructure and macrostructure of image patterns. We also provide several dissimilarity measures based on MSOF for object recognition. Finally we apply MSOF for face recognition. The experiments on FERET database illustrate that our proposed method achieves an excellent performance. We believe that the success of SOF is not limited to faces. Since SOF is general for appearance-based object representation, our future work will be applying SOF on other object classification or recognition problems to investigate its power.

Acknowledgements. This work was supported by the following funding resources: National Natural Science Foundation Project #60518002, National Science and Technology Supporting Platform Project #2006BAK08B06, National 863 Program Projects

#2006AA01Z192 and #2006AA01Z193, Chinese Academy of Sciences 100-people project, and the AuthenMetric Research Foundation.

References

1. Ahonen, T., Hadid, A., Pietikainen, M.: Face recognition with local binary patterns. In: Proceedings of the European Conference on Computer Vision, Prague, Czech, pp. 469–481 (2004)
2. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 711–720 (1997)
3. Crow, F.: Summed-area tables for texture mapping. In: SIGGRAPH, vol. 18(3), pp. 207–212 (1984)
4. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: a statistical view of boosting. Technical report, Department of Statistics, Sequoia Hall, Stanford Univerity (July 1998)
5. Lades, M., Vorbruggen, J., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R.P., Konen, W.: Distortion invariant object recognition in the dynamic link architecture. IEEE Transactions on Computers 42, 300–311 (1993)
6. Liao, S., Lei, Z., Zhu, X., Sun, Z., Li, S., Tan, T.: Face recognition using ordinal features, pp. 40–46 (2006)
7. Liu, C., Wechsler, H.: Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. IEEE Transactions on Image Processing 11(4), 467–476 (2002)
8. Moghaddam, B., Nastar, C., Pentland, A.: A Bayesain Similarity measure for direct image matching. Media Lab Tech. Report No.393, MIT (August 1996)
9. Ojala, T., Pietikainen, M., Harwood, D.: A comparative study of texture measures with classification based on feature distributions. Pattern Recognition 29(1), 51–59 (1996)
10. Penev, P., Atick, J.: Local feature analysis: A general statistical theory for object representation. Neural Systems 7(3), 477–500 (1996)
11. Sadr, J., Mukherjee, S., Thoresz, K., Sinha, P.: 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)
12. Schapire, R.E., Singer, Y.: Improved boosting algorithms using confidence-rated predictions. In: Proceedings of the Eleventh Annual Conference on Computational Learning Theory, pp. 80–91 (1998)
13. Schneiderman, H.: Toward feature-centric evaluation for efficient cascaded object detection. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, pp. 1007–1013 (June 27 - July 2, 2004)
14. Simard, P.Y., Bottou, L., Haffner, P., Cun, Y.L.: Boxlets: a fast convolution algorithm for signal processing and neural networks. In: Kearns, M., Solla, S., Cohn, D. (eds.) Advances in Neural Information Processing Systems, vol. 11, pp. 571–577. MIT Press, Cambridge (1998)
15. Sinha, P.: Toward qualitative representations for recognition. In: Proceedings of the Second International Workshop on Biologically Motivated Computer Vision, Tubingen, Germany, pp. 249–262 (November 22-24, 2002)
16. Thoresz, K.J.: On qualitative representations for recognition. Master's thesis, MIT (July 2002)
17. Turk, M.A., Pentland, A.P.: Eigenfaces for recognition. Journal of Cognitive Neuroscience 3(1), 71–86 (1991)
18. Viola, P., Jones, M.: Robust real time object detection. In: IEEE ICCV Workshop on Statistical and Computational Theories of Vision, Vancouver, Canada (July 13, 2001)
19. Wiskott, L., Fellous, J., Kruger, N., Malsburg, C.v.d.: Face recognition by elastic bunch graph matching. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 775–779 (1997)