FACE RECOGNITION BASED ON SIGMA SETS OF IMAGE FEATURES

Ramya Srinivasan [1], Abhishek Nagar [2], Anshuman Tewari [2], Donato Mitrani [2], Amit Roy-Chowdhury [1]*

[1] Electrical Engineering Department, University of California, Riverside, [2] Samsung Research America, Dallas

ABSTRACT

Automatic face recognition is prevalent in a wide range of systems these days and it is critical to explore new techniques in order to enhance the state of the art. In this paper, we analyze the Region Covariance Matrix (RCM) and its enhancement based on Sigma sets as a feature extraction procedure for face images. The RCM features encode the covariance of various low level features,e.g., pixel intensities and gradients. Sigma sets, on the other hand, reduce the computational complexity of comparing two RCMs. Based on our experiments on the Labeled Faces in the Wild (LFW) dataset, we show that the proposed technique outperforms the popular Local Binary Patterns (LBP) technique and is on par with other better performing techniques that use complex classifiers.

Index Terms— Face recognition, Region covariance matrix, Sigma sets.

1. INTRODUCTION

Face recognition is a form of biometric identification [12] involving recognition of individuals based on the salient characteristics of their face images. Be it for government use such as law enforcement, voter identification, surveillance and immigration, or for commercial use such as gaming industry, face tagging on internet, e-commerce, healthcare and banking, a large number of real world applications utilize face recognition. As a result, there has been enormous interest in this area of research.

A variety of challenges are associated with a typical face recognition task. While some may involve accounting for facial aging, marks, and facial expressions, others may be due to non-rigid motion and background clutter, typical in videos; yet some others may involve comparing face images over different media such as a sketch to a photograph or a near infared image to a photo. A critical survey of still image and video based face recognition research is provided in [13].

It has been observed that the performance of several state-of-the-art face recognition methods degrades to a large extent in unconstrained environments. Figure 1 illustrates some cases where a common face recognition algorithm based on Local Binary Patterns (LBP) [28] fails to identify the image pairs belonging to same person (shown in same column) as match pairs. One effective approach to alleviate these limitations is by designing highly discriminative, robust and yet simple image features that can be efficiently matched.

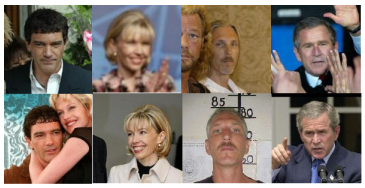

Fig. 1. Example image pairs from the LFW where LBP fails. These instances illustrate some typical sources of variations mentioned in text: poor image quality, occlusions, and difference in facial expression. Images in the same column correspond to match pairs.

A typical face recognition system consists of two main stages, namely, a feature extraction stage followed by a matching stage. The first stage primarily involves extracting low or high level informative features either manually or automatically from images or videos. Depending on the representation in feature space, these approaches can be broadly divided into three categories, namely, holistic methods which use global image information, such as [14], feature based structural matching methods, e.g., [15] and hybrid methods involving both holistic and structural matching procedures such as [16]. While holistic methods are often simple and computationally efficient, feature based structural matching methods are more robust than holistic methods but often require complex matching procedure.

The matching stage involves classification of an input image into one of the pre-determined classes. In case of one-to-one matching, the classes being considered are "match" and "non-match", and in case of one-to-many matching the different classes indicate different identities registered in the system. The classifiers used in this stage are typically learned offline using training data, in an unsupervised [1], semi-supervised [2] or supervised framework [3].

The quality of features, which are extracted based on expert knowledge [17] and/or learned from available training data [18], is of critical importance. Among the holistic features, the low level ones such as color, gradient and filter responses are the simplest choices [4, 5], but as such they are not robust in the presence of illumination changes and non-rigid motion and also typically have high dimensionality. In

*The work was performed while the author was interning at Samsung.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

Fig. 2. Performance of sigma set descriptors on the same set of image pairs considered in Figure 1. Blue box indicates correct result while red denotes incorrect result.

this paper, we explore the use of Region Covariance Matrices (RCM) as the set of holistic face image features and propose further improvements in terms of feature representation and matching procedure using sigma sets.

Region Covariance Matrices: The region covariance matrix (RCM) of a local region in an image with respect to a set of low level features essentially consists of the local 2^{nd} order statistics of the given features. RCM was first introduced as a region descriptor by Tuzel et al. [6] and has since been applied to many applications such as texture classification, object detection [6, 19], object tracking [7, 20] and human detection [8]. A major advantage of the covariance matrices is that they have much lower dimensionality when compared to the existing low level image features and can be computed efficiently using integral images [6]. RCMs are robust to small pose variations and provide a natural way for fusing multiple features which might be correlated. These descriptors have shown best discriminative power for human detection tasks [9] and have also been recently studied in the context of face recognition [10].

However, the main drawback of RCMs is that these descriptors do not lie in a Euclidean space. For example, the RCMs are not closed under multiplication with negative scalars. In order to accommodate this fact, the arithmetic operations on RCMs are borrowed from Riemmannian geometry which are usually computationally complex.

Sigma Sets: Recently, Hong et al. [11] proposed a region descriptor called sigma sets that possesses the effectiveness of 2^{nd} order statistics while at the same time has a highly efficient matching procedure compared to RCM. Sigma set consists of a small set of vectors that can be uniquely determined through Choloesky decomposition on the covariance matrix. Further, distance between two sigma point-sets can be efficiently computed using an approximation of Hausdorff distance metric. Of late, sigma sets have been employed for various applications [21], although we are not aware of any work that has explored these descriptors for face recognition.

Motivated by the aforementioned attributes of sigma set descriptors, we study their feasibility in the context of face identification in the wild in this work. An illustration of the benefits offered by these descriptors can be seen in Figure 2. Image pairs enclosed in blue are the ones correctly identified as "match pairs" by sigma set descriptors while the one enclosed in red denotes an incorrect result. This is substantially better than LBP.

The rest of the paper is organized as follows. In Section 2, we describe the proposed approach including the specific details of RCM and Sigma sets construction. In Section 3, we describe the details of the experimental framework and provide a comparison of the proposed technique with other state of the art procedures. Section 4 provides the conclusions.

2. PROPOSED APPROACH

In the proposed approach, first, all face images are cropped to a size of 200×200. Then, a set of low level features are extracted from the face image. For this, we mainly consider filters that capture local averaging and local gradient. These include Gaussian, Sobel and Gabor filters. It is to be noted that no pose correction is done on these images. Region covariance matrices and subsequently sigma sets are then constructed from these features. These form the feature descriptors for the images under consideration. During matching stage, Manhattan distance between sigma sets is used as the distance metric.

We first describe the features used in constructing sigma sets below.

Gaussian filter: Gaussian filters highlight average brightness in a local region where the size of the locality is determined by the scale or the variance of the 2D Gaussian filter kernel used. A Gaussian kernel is described as

$$f(x,y) = Ae^{\left(-\left(\frac{(x-x_0)^2}{2\sigma_x^2}\right)+\left(\frac{(y-y_0)^2}{2\sigma_y^2}\right)\right)}, \quad (1)$$

where best scale for our experiments was determined by varying the values values of σ_x and σ_y in steps of 0.1. Here x_0 and y_0 determine the center of the filter.

Sobel filter: Sobel filters highlight the change in intensity in a local region. We convolve the image with horizontal and vertical Sobel kernels to capture the complete gradient information.

Gabor filter: Gabor filters aggregate directional gradients and are able to effectively capture facial features. A 2D Gabor kernel is the product of a Gaussian and a sine or cosine plane wave. A 2D Gabor filter can be described as

$$g(x,y) = Ke^{-\pi(a^2(x-x_0)^2+b^2(y-y_0)^2)}e^{j2\pi(u_0x+v_0y)}, \quad (2)$$

where a and b determine the scale of the filter and u_0 and v_0 determine the frequency component of the filter. The particular value of scale is chosen after testing performance on five standard scales, retaining the best one. For our analysis, we consider eight different sine wave orientations and five different combinations of frequencies and Gaussian scales.

After the filter responses are obtained, the image is uniformly divided into 8×8 equal sized blocks and region covariance matrices (RCM) based on the above mentioned features are computed for each block.

2.1. Covariance Descriptor Computation

Let I denote a face image. For a given rectangular region $R \in I$ with N pixels, let $\vec{f_i}$ be a d dimensional feature vector (consisting of Gabor, Gaussian and/or gradient responses) extracted from the i^{th} pixel in R , and $\vec{u}$ be the mean vector of the set of feature vectors $\vec{f_i}$ in R. The $d \times d$ covariance matrix $C(R)$ of R can be calculated as

$$C(R) = F_R F_R^T \tag{3}$$

where $F_R = [\hat{f}_1, ...\hat{f_N}]$ denotes the matrix of centered vectors $\hat{f}_i = \frac{1}{\sqrt{(N)}}(\vec{f_i} - \vec{u})$.

2.2. Sigma Sets Computation

A Sigma set associated with a covariance matrix is a small canonical set of points S that have same covariance values as the given matrix. In other words, the set of points S in a Sigma set is *equivalent* to the set of feature vectors obtained from the region R in terms of 2^{nd} order statistics. Mathematically, it follows from Eq. 3 that for any matrix L that satisfies $C(R) = LL^T$, the set of columns of L has the same second order statistic as R. One way to obtain such a decomposition of the covariance matrix is using Cholesky decomposition. A Cholesky decomposition is used to represent any symmetric positive-definite matrix, such as a covariance matrix, as a product of a lower triangular matrix and its transpose. The fact that the component matrix of a Cholesky decomposition is lower triangular, it imposes a canonical order on the set of points it represents and this is very helpful in devising a simple distance metric between two sigma sets.

The sigma set computation can be summarized algorithmically as follows:

Given: A face region R consisting of N, $d \times 1$ feature vectors.
Output: Sigma set $S = [L_1, ..., L_d]$ satisfying $C(S) = C(R)$.
Algorithm:
1. Calculate the $d \times d$ covariance matrix $C = C(R)$ of the face region R.
2. Perform Cholesky decomposition of C, $C = LL^T$, where L is a lower triangular matrix.
3. Multiply L by the scalar $\sqrt{d}$, i.e., $L = \sqrt{d} \times L$.
4. $S = [L_1, ..., L_d]$ where L_i is the i^{th} column of L.

2.2.1. Distance measure

The distance between sigma sets can be evaluated as summation of point to point distance and is equivalent to modified Hausdroff distance (MHD) [26], a widely used distance metric over closed and bounded sets. Given two sigma sets S_A and S_B, the modified Hausdroff distance is defined as

$$H(S_A, S_B) = \max\{h(S_A, S_B), h(S_B, S_A)\}, \tag{4}$$

where $h(S_A, S_B)$ is the forward distance measure obtained by matching the points in S_A to points in S_B and $h(S_B, S_A)$ is the backward distance that corresponds to matching S_B to S_A. Here,

$$h(S_A, S_B) = \frac{1}{d}\sum_{i=1}^{d} \min_{j=1}^{d}(d_E(L_i^A, L_j^B)), \tag{5}$$

where L_i^A and $L_j^B, i = 1, ...d$ denote the i^{th} and j^{th} points in S_A and S_B respectively and d_E is a preferred distance metric; we use Manhattan distance in our experiments..

Since the structure of the sigma set enforces the first i elements of i^{th} sigma point to be zero, we can assume that

$$d_E(L_i^A, L_j^B) = \infty, i \neq j. \tag{6}$$

The distance $h(S_A, S_B)$ is thus given by

$$h(S_A, S_B) = \frac{1}{d}\sum_{i=1}^{d} d_E(L_i^A, L_i^B). \tag{7}$$

Since $d_E(L_i^A, L_i^B) = d_E(L_i^B, L_i^A)$ thus $h(S_A, S_B) = h(S_B, S_A)$ and the Eq. (4) becomes

$$H(S_A, S_B) = \frac{1}{d}\sum_{i=1}^{d} d_E(L_i^A, L_i^B). \tag{8}$$

This distance measure is then used to accept or reject a matching hypothesis based on certain system threshold.

3. EXPERIMENTS

Dataset: We conduct experiments on the Labeled Faces in the Wild (LFW) data-set and use the image restricted configuration. The data-set consists of 13,233 images of 5,749 people, which are organized into 2 views – a development set of 2,200 pairs for training, 1,000 pairs for testing (model selection), and a 10-fold cross-validation set of 6,000 pairs, on which to evaluate final performance using the average pairwise accuracy metric as described in [22].

3.1. Performance Evaluation

In computing the distance between corresponding sigma sets of two faces, we consider the four following possibilities:
1. Performance using a single scale Gabor filter, Gaussian and Gradient Features (3G):
First, we consider single scale Gabor filter responses across 8 different orientations along with 4 Gaussian and 2 Sobel responses for constructing sigma sets. Average pairwise matching accuracy on the cross-validation set was 79.54%.
2. Block-wise weighted distance between sigma sets 3G: Since different regions of the face contribute in different proportions for identification purposes, we next consider weighting the distance values computed for different regions by their

precision. The distance between two images I_1 and I_2 in this scenario is computed as

$$D(I_1, I_2) = \frac{1}{d} \sum_{k=1}^{T} w_k \sum_{i=1}^{d} d_E(L_i^{b_k}, L_i^{b_k}), \qquad (9)$$

where w_k is the learned weight (pairwise accuracy for corresponding blocks as learned from training instances) for the corresponding blocks b_k in two face images and T is the total number of blocks. The average accuracy in this scenario was 82.32%.

3. Block-wise scale selection for Gabor filter along with Gaussian and Gradient Features: Here, we learn the performance of five different scales for individual regions and finally use the scale that yields the best performance for individual regions. The selected Gabor responses are then combined with Gaussians and Gradient responses to evaluate performance. The best kernel size for Gaussian filter is determined experimentally from the training data. With this setup, average accuracy across the 10 folds was 81.73%.

4. Block wise scale selection for Gabor filter along with Gaussian and Gradient features using the weighted distance measure: Finally, we incorporated block wise scale selection for the Gabor filters along with Gaussian and Gradient features and used the weighted distance measure to compute distance between the constructed 3G sigma sets. This resulted in an average accuracy of 83.03%.

The corresponding ROC curves are depicted in Figure 4.

Computational Complexity: The amount of computations

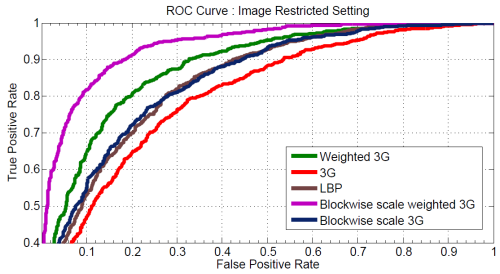

Fig. 3. ROC curves for the various scenarios.

required for Cholesky decomposition (CD) of a $d \times d$ matrix is $O(d^3)$ which is relatively small given that there are relatively few features being processed. Additionally, CD ensures fast distance computation that is linear in terms of the number of feature elements unlike for RCM where one would require complex computation of matrix expoenential or logarithm.

Comparison with other methods: Based on the characteristics of our technique, we mainly focus on the image restricted scenario where no additional training data is allowed. In this scenario, the proposed technique is 1.4% more accurate than the LBP features of the same dimensions (as that used for sigma sets) which produce an accuracy of 81.65%. In fact, the dimensions of LBP features which produce one of the best results on LFW dataset [29] is roughly 8 times more than our feature dimensions. Also, the dimension of our features is $1/5^{th}$ in comparison to [24] thereby significantly reducing the computational cost.

Method	Accuracy in %	Comments
[25]	79.08 ± 0.14	aligned images
[27]	79.35 ± 0.55	aligned images
Sigma sets	83.03 ± 0.49	non-aligned images, no training on view 2
[23]	84.08 ± 1.20	aligned images
[24]	87.47 ± 1.49	aligned images, training on view 2.

Table 1. Comparison of the proposed method with state-of-the-art.

Method	Good	Bad	Ugly
LBP	84.2	80.2	77.2
Sigma sets	86.1	83.5	79.3

Table 2. Quality based performance of the proposed features.

Moreover, unlike [24], we do not perform training the model on 9 splits and testing on 10^{th} split in view 2; instead we directly test on each of the 10 splits using the parameters learned from view 1. Despite the fact that the techniques proposed in [24, 23] use trained classifiers (on view 2) and also leverage upon the pre-registration of images, our method yields comparable matching performance. It is to be noted that the proposed method can be extended to incorporate outside training data which can further improve the performance. Table 1 provides a comparison of the proposed method against some state-of-the-art techniques under image restricted scenario.

Quality based performance: Here, we divide the set of images, based on the maximum matching scores obtained from them while comparing with other images including true matches, into good (lowest distance scores), bad and ugly (highest distance scores) categories such that each category has same number of images. This division was based on sigma sets-derived distance. We then compare the accuracy of the proposed technique with LBP in each of these categories. As shown in Table 2, the proposed method performs consistently better than LBP for images across all qualities.

4. CONCLUSIONS

We analyzed the feasibility of a new set of face descriptors called sigma sets constructed from simple image features. Experiments show promising performance on the challenging LFW database. Future work will involve in testing these descriptors on video sequences. We also plan to explore different feature selection processes for constructing sigma sets.

5. REFERENCES

[1] B. Raytchev, and H. Murase. Unsupervised Face Recognition by Associative Chaining. *Pattern Recognition*, 2003.

[2] Z. Cui, W. Li, D. Zhu, S Shan, and X. Chen. Fusing Robust Face Region Descriptors via Multiple Metric Learning for Face Recognition in the Wild. *IEEE Intl. Conf. on Computer Vision and Pattern Recognition*, 2013.

[3] D. Rim, K. Hassan, and C. Pal. Semi Supervised Learning in Wild Faces and Videos. *British Machine Vision Conference*, 2011.

[4] A. Rosenfeld, and G. Vanderburg. Coarse-fine Template Matching. *IEEE. Trans. Syst. Man. Cyb.*,1977.

[5] R. Brunelli, and T. Poggio. Face Recognition: Features versus Templates. *IEEE. Trans. Patt. Anal. Machine. Intell.*, 1993.

[6] O. Tuzel, F. Porikli, and P. Meer. Region Covariance: A Fast Descriptor for Detection and Classification. *European Conference on Computer Vision.*, 2006.

[7] O. Tuzel, F. Porikli, and P. Meer. Covariance Tracking Using Model Update Based on Lie Algebra. *IEEE. Conf. on Computer Vision and Pattern Recognition*, 2006.

[8] O. Tuzel, F. Porikli, and P. Meer. Pedestrian Detection via Classification on Riemmanian Manifolds. *IEEE. Trans. Patt. Anal. Machine. Intell.*, 2008.

[9] B. Wu, and R. Nevatia. Optimizing Discrimination Efficiency Tradeoff in Integrating Heterogenous Local Features for Object Detection. *IEEE. Conf. on Computer Vision and Pattern Recognition*, 2008.

[10] Y. Pang, T. Yuan and X. Li. Gabor-Based Region Covariance Matrices for Face Recognition. *IEEE. Transactions on Circuits and Systems for Video Technology*, 2008.

[11] X. Hong, T. Yuan, H. Chang, S. Shan, X. Chen and W. Gao. Sigma Set: A Small Second Order Statistical Region Descriptor. *IEEE.Conf. on Computer Vision and Pattern Recognition*, 2009.

[12] A. K. Jain, A. Ross, and K. Nandakumar. Introduction to Biometrics: A textbook. *Springer Publishers ISBN 978-0-387-77325-4*, 2011.

[13] W. Zhao, and R. Chellappa. Face Processing: Advanced Modeling and Methods. *Academic Press*,2006

[14] M. Turk, and A. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 1991.

[15] Wiskott et al. Face Recognition by Elastic Graph Bunch Matching. *IEEE. Trans. Patt. Anal. Mach. Intell.*,1997.

[16] A. Pentland, B. Moghaddam, and T. Starner. View Based and Modular Eigenspaces for Face Recognition. *IEEE. Conf. Comp. Vision and Patt. Reco.*, 1994.

[17] C. Liu, and H. Wechsler. Gabor Feature Based Classification Using the Enhanced Fischer Linear Discriminant Model for Face Recognition. *IEEE. Trans. on Image Proc.*, 2002.

[18] Z. Cao, Q. Yin , X. Tang, and J. Sun. Face Recognition with Learning Based Descriptors. = *IEEE. Conf. Comp. Vision and Patt. Reco.*, 2010.

[19] F. Porikli and T. Kocak. Robust License Plate Detection Using Covariance Matrices in a Neural Network Framework. *IEEE.Adv. Video Signal Based Surveillance Conf.*, 2006.

[20] F. Porikli and T. Kocak. Fast Construction of Covariance Matrices for Arbitrary Size Image Windows. *IEEE.Intl. Image Proc. Conf.*, 2006.

[21] E. Erdem and A. Erdem. Visual Saliency Estimation by Nonlinearly Integrating Features Using Region Covariances. *Journal of Vision*, 2013.

[22] G. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. *European Conference on Computer Vision Workshop*, 2008 .

[23] H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang. Probabilistic Elastic Matching for Pose Variant Face Verification. *IEEE. Conf. on Computer Vision and Pattern Recognition*, 2013 .

[24] K. Simonyan, O. Parkhi, A. Vedaldi, and A. Zisserman. Fisher Vector Faces in the Wild. *British Machine Vision Conference*, 2013 .

[25] S. Arashloo,and J. -Kittler. Efficient Processing of MRFs for Unconstrained- Pose Face Recognition. *Biometrics Theory, Applications and Systems*, 2013 .

[26] M. Dubuisson and A. Jain. A Modified Hausdroff Distance for Object Matching. *International Conference on Pattern Recognition*, 1994.

[27] N. Pinto, J. . DiCarlo, and D. Cox. How far can you get with a modern face recognition test set using only simple features? *IEEE Conf. on Computer Vision and Pattern Recognition*, 2009.

[28] T. Ojala, M. Pietikinen, and D. Harwood. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. *Proceedings of the 12th IAPR International Conference on Pattern Recognition*, 1994.

[29] D. Chen, X. Cao, F. Wen, and J. Sun. Blessing of Dimensionality: High-dimensional Feature and Its Efcient Compression for Face Verication. *IEEE International Conference on Computer Vision and Pattern Recognition*, 2013.