SEMI-SUPERVISED DIMENSIONALITY REDUCTION ON DATA WITH MULTIPLE REPRESENTATIONS FOR LABEL PROPAGATION ON FACIAL IMAGES

Olga Zoidi, Nikos Nikolaidis, Ioannis Pitas

Department of Informatics
Aristotle University of Thessaloniki
Box 451, Thessaloniki 54124, GREECE
tel: +30 2310 996361
{ozoidi, nikolaid, pitas}@aiia.csd.auth.gr

ABSTRACT

In this paper a novel method is introduced for semi-supervised dimensionality reduction on facial images extracted from stereo videos. It operates on image data with multiple representations and calculates a projection matrix that preserves locality information and a priori pairwise information, in the form of must-link and cannot-link constraints between the various data representations, as well as label information for a percentage of the data. The final data representation is a linear combination of the projections of all data representations. The performance of the proposed Semi-supervised Multiple Locality Preserving Projections method was evaluated in person identity label propagation on facial images extracted from stereo movies. Experimental results showed that the proposed method outperforms state of the art methods.

Index Terms— Locality preserving projections, semi-supervised learning, label propagation

1. INTRODUCTION

Dimensionality reduction refers to the procedure of projecting high-dimensional data onto a subspace of the original high-dimensional space. When the data are M facial images of size $N_x \times N_y$ pixels, a typical dimensionality reduction method is to find a projection matrix $\mathbf{A} \in \Re^{N \times L}$ that maps the images $\mathbf{x}_i \in \Re^N$, $i = 1, \dots, M$, $N = N_x N_y$ on a subspace $\Re^L$, $L << N$. The rows of the projection matrix $\mathbf{A}$ form the basis vectors in the resulting space $\Re^L$ and the data projections $\mathbf{x}'_i = \mathbf{A}^T \mathbf{x}_i \in \Re^L$ form the facial image features to be used in the classification task. Such a popular subspace representation widely used in person recognition algorithms, that operates on graphs, is the Locality Preserving Projection (LPP) [1–3]. The objective of LPP is the projection of the original high-dimensional data to a reduced dimensional space, so that the projected data inherit the locality information of the original data, i.e, when $\|\mathbf{x}_i - \mathbf{x}_j\|_2$ is small, then $\|\mathbf{x}'_i - \mathbf{x}'_j\|_2$ is small as well. Sparsity constraints in the objective function of LPP were imposed in [4], so that the sparse reconstructive weights are preserved, while in [3], a regularized LPP method was presented, that extracts useful discriminant information from the entire feature space. Moreover, orthogonality constraints were imposed on the discriminant LPP in [5]. LPP performs unsupervised dimensionality reduction, because the only information it exploits is from the data structure in the original space $\Re^N$. Several extensions of LPP have been proposed in the literature, that incorporate prior information about the data and extend LPP to the semi-supervised and the fully-supervised framework, such as [6–8]. In [6], LPP aims, apart from locality preservation, the maximization of the between-class distance and the minimization of the within-class distance of a small number of available labeled data, while [7] maximizes the difference between the locality preserving between-class scatter matrix and locality preserving within-class scatter matrix. Finally, [8] exploits a priori pairwise constraints on the original data.

In this paper, we propose a novel method for semi-supervised dimensionality reduction on data with multiple representations that finds application in person identity label propagation on stereo videos. Generally the multiple data representations are obtained when multiple features are employed for data description. In the case of stereo facial images, the multiple representations used are the left and right channel facial images. The method exploits information obtained from multiple data representations, by finding a projection matrix that preserves locality information and additional a priori information between the data in all data representations. The proposed method searches for a projection matrix that projects all data representations on the same reduced-dimensional space. Each data representation influences the projection matrix with a weight that is automatically learned from the regularization framework. The

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement number 287674 (3DTVS). This publication reflects only the author's views. The European Union is not liable for any use that may be made of the information contained therein.

projections of all data representations are then combined into a single reduced-dimensional representation, which will be used in the classification task. Experimental results on person identity label propagation on stereo videos showed that the proposed dimensionality reduction framework outperforms state of the art dimensionality reduction methods based on LPP.

2. SIMILARITY GRAPH CONSTRUCTION

Let $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ be the set of M data that belong to classes $\mathcal{C} = \{1, \dots, C\}$. Each sample belongs only in one class. The data have K representations. For the k-th representation method, the data matrix $\mathbf{X}_k = [\mathbf{x}_{1,k}, \dots, \mathbf{x}_{M,k}] \in \Re^{N \times M}$ is constructed, where $\mathbf{x}_{i,k}$ denotes the feature vector of the k-th representation of the i-th sample. Let us consider that the class labels $l(\mathbf{x}_j) \in \mathcal{C}$, $j = 1, \dots, n_l$ of n_l data are known. Let $\mathcal{G} = (\mathcal{X}, \mathcal{E})$ be the graph, whose nodes are the data entries $\mathbf{x}_i$ in the set $\mathcal{X}$ and whose edges are the pairwise data relationships. The edge in the graph that connects the nodes i and j is assigned with a value W_{ij} that indicates the similarity between the adjacent graph nodes. Usually, this similarity is computed according to the heat kernel equation [1]:

$$W_{ij} = e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma}}, \tag{1}$$

were σ is the mean edge length distance among neighbors. In the proposed method, we incorporate the label information into the similarity matrix $\mathbf{W}'$ as follows:

$$W'_{ij} = \begin{cases} 1, & \text{if } l(\mathbf{x}_i) = l(\mathbf{x}_j) \\ 0, & \text{if } l(\mathbf{x}_i) \neq l(\mathbf{x}_j) \\ W_{ij}, & \text{if } \nexists \mathbf{x}_k \in \mathcal{N}_i | l(\mathbf{x}_k) \neq l(\mathbf{x}_i) \& W_{kj} > W_{ij} \\ aW_{ij}, & \text{if } \exists \mathbf{x}_k \in \mathcal{N}_i | l(\mathbf{x}_k) \neq l(\mathbf{x}_i) \& W_{kj} > W_{ij} \end{cases}, \tag{2}$$

where $l(\mathbf{x}_i)$ and $\mathcal{N}_i$ denote the label and the neighborhood of sample $\mathbf{x}_i$, respectively, and $0 \leq a \leq 1$ is a penalty parameter. The first two cases in (2) refer to the pairwise similarity between the labeled samples i and j and the last two refer to the pairwise similarity between the labeled sample i and the unlabeled sample j. In experiments we set $a = 0.5$. One similarity matrix $\mathbf{W}_k$ is constructed for each data representation.

3. PAIRWISE CONSTRAINTS

Let $\mathcal{S}$ be the set of similar pairs:

$$\mathcal{S} = \{(i,j) | \mathbf{x}_i, \mathbf{x}_j \text{ must have the same label}\}, \tag{3}$$

and $\mathcal{D}$ be the set of dissimilar pairs:

$$\mathcal{D} = \{(i,j) | \mathbf{x}_i, \mathbf{x}_j \text{ must have different labels}\}. \tag{4}$$

Two weight matrices are constructed, $\mathbf{W}_s$ and $\mathbf{W}_d$, for the similar and dissimilar constraints, respectively, as follows:

$$W_{s,ij} = \begin{cases} 1, & \text{if } (i,j) \in \mathcal{S} \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

$$W_{d,ij} = \begin{cases} 1, & \text{if } (i,j) \in \mathcal{D} \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

Intuitively, if we know that two nodes have the same labels from prior knowledge, then the neighbors of these nodes should also have the same label, due to neighboring node similarity. In a similar argumentation, if we know that two nodes have dissimilar labels, then the nodes that belong to the neighborhood of one node should have different label from the other node and vice versa. This means that we can generalize the pairwise constraints to include neighboring nodes in an iterative procedure, similarly to label propagation. Let $\mathcal{N}_i$ be the neighborhood of node i, based on, e.g., thresholding the Euclidean distance between two nodes and $\mathbf{P} \in \Re^{M \times M}$ be a sparse weight matrix with entries:

$$P_{ij} = \begin{cases} \frac{1}{|\mathcal{N}_i|}, & \text{if } j \in \mathcal{N}_i \\ 0, & \text{otherwise,} \end{cases} \tag{7}$$

where $|\mathcal{N}_i|$ is the cardinality of the set $\mathcal{N}_i$. It is clear that the sum of each row of $\mathbf{P}$ is 1. We define a function $\mathbf{F}_s$ that assigns a real value to every graph node that indicates its label similarity to the other graph nodes. In each iteration, the node incorporates some information from its neighbors and retains some information from its initial state $\mathbf{W}_s$. At t-th iteration, the label similarity is equal to:

$$\mathbf{F}_s^{(t)} = a\mathbf{P}\mathbf{F}_s^{(t-1)} + (1-a)\mathbf{W}_s, \tag{8}$$

which converges to the steady state [9]:

$$\mathbf{F}_s = (1-a)(\mathbf{I} - a\mathbf{P})^{-1}\mathbf{W}_s. \tag{9}$$

Similarly, the label dissimilarity is propagated according to:

$$\mathbf{F}_d^{(t)} = a\mathbf{P}\mathbf{F}_d^{(t-1)} + (1-a)\mathbf{W}_d, \tag{10}$$

which converges to the steady state:

$$\mathbf{F}_d = (1-a)(\mathbf{I} - a\mathbf{P})^{-1}\mathbf{W}_d. \tag{11}$$

4. SEMI-SUPERVISED LPP ON MULTIPLE GRAPHS

In this paper, we propose a novel method for performing linear dimensionality reduction on data with multiple representations, where the class (label) of n_l data is known. The mean class vectors of the labeled samples are computed by:

$$\mathbf{m}_c = \frac{1}{|\mathcal{L}_c|} \sum_{\mathbf{x}_i \in \mathcal{L}_c} \mathbf{x}_i, \quad c \in \mathcal{C}, \tag{12}$$

where $\mathcal{L}_c$ is the set of samples with label $c \in \mathcal{C}$. The mean class vectors similarity matrix is given by the heat kernel equation:

$$B_{ij} = e^{-\frac{\|\mathbf{m}_{c_i} - \mathbf{m}_{c_j}\|^2}{\sigma}}. \tag{13}$$

The proposed method takes into account locality information, obtained from the data similarity matrix, class information, obtained from the labeled data, and a priori pairwise similarity and dissimilarity constraints. It searches for a $N \times L$ projection matrix $\mathbf{A}$ that operates on all visual data views (e.g., the left/right video channel) and also searches for the optimal linear combination of the data projections. Let $\mathbf{X}_k$, $k = 1, \ldots, K$ be the different representations data matrix. The objective of the proposed method is the minimization of the function:

$$\arg\min_{\mathbf{A},\boldsymbol{\tau}} \sum_{k=1}^{K} \tau_k \left[\mathrm{tr}\left(\mathbf{A}^T\mathbf{X}_k\mathbf{L}_{W_k}\mathbf{X}_k^T\mathbf{A}\right) + \gamma \mathrm{tr}\left(\mathbf{A}^T\mathbf{X}_k\mathbf{L}_S\mathbf{X}_k^T\mathbf{A}\right) - \delta \mathrm{tr}\left(\mathbf{A}^T\mathbf{X}_k\mathbf{L}_D\mathbf{X}_k^T\mathbf{A}\right) - \zeta\left(\mathbf{A}^T\mathbf{M}_k\mathbf{L}_{Bk}\mathbf{M}_k^T\mathbf{A}\right)\right] + \varepsilon\|\boldsymbol{\tau}\|^2 \quad (14)$$

subject to the constraints:

$$\mathbf{A}^T\mathbf{A} = \mathbf{I}, \ \ \sum_{k=1}^{K} \tau_i = 1, \ \ \tau_i \geq 0, \quad (15)$$

where γ, δ, ζ are parameters that regulate the significance of the pairwise similarity and dissimilarity constraints and the mean class vectors similarities, respectively and ε is a regularization parameter that punishes the coefficients vector $\boldsymbol{\tau}$ to take increased value for only one image representation. The first constraint in (15) ensures that the projection matrix $\mathbf{A}$ is orthonormal. The matrices $\mathbf{L}_{Wk}$, $\mathbf{L}_S$, $\mathbf{L}_D$ and $\mathbf{L}_{Bk}$ are the graph Laplacians defined by the weight matrices $\mathbf{W}'_k$, $\mathbf{F}_s$, $\mathbf{F}_d$ and $\mathbf{B}_k$, respectively, according to:

$$\mathbf{L} = \mathbf{D}^{-1/2}(\mathbf{D} - \mathbf{W})\mathbf{D}^{-1/2}, \quad (16)$$

where $\mathbf{D}$ is a diagonal matrix with elements $D_{ii} = \sum_j W_{ij}$. $\mathbf{L}_{Wk}$, $\mathbf{L}_{Bk}$ vary according to the data representation, while $\mathbf{L}_s$, $\mathbf{L}_d$ are constant for all representations. Finally, the matrices $\mathbf{M}_k \in \Re^{N\times C}$ the matrix with columns the mean class vectors for the k-th representation. The first trace in (14) ensures that the locality information of the data in the original high dimensional space is preserved in the projected space. The second/third trace in (14) ensures that similar/dissimilar data pairs are mapped close to/away from each other. Finally, the fourth trace in (14) forces the classes mean vectors to be mapped away from each other.

By selecting the parameters γ, δ and ζ so that the matrix $\mathbf{X}_k\left(\mathbf{L}_{Wk} + \gamma\mathbf{L}_s - \delta\mathbf{L}_d\right)\mathbf{X}_k^T - \zeta\mathbf{M}_k\mathbf{L}_{Bk}\mathbf{M}_k^T$ is positive semi-definite, the cost function (14) under the constraints (15) is convex, with respect to the variables $\mathbf{A}$ and $\boldsymbol{\tau}$. In practice, the value of γ is chosen to be larger than the values of δ and ζ by several orders of magnitude. In the experiments, we set $\gamma = 100$ and $\delta = \zeta = 0.01$. The optimization problem is solved iteratively for $\mathbf{A}$ and $\boldsymbol{\tau}$ as follows:

1. First, $\boldsymbol{\tau}$ is initialized with the values $\tau_k = \frac{1}{K}$, $k = 1, \ldots, K$.
2. The system (14), (15) is solved for $\mathbf{A}$ by solving the following eigenvalue problem:

$$\left[\sum_k \tau_k\mathbf{X}_k\left(\mathbf{L}_k + \beta\mathbf{L}_s - \gamma\mathbf{L}_d\right)\mathbf{X}_k^T - \sum_k \zeta\mathbf{M}_k\mathbf{L}_{Bk}\mathbf{M}_k^T\right]\mathbf{a}_l = \lambda\mathbf{a}_l. \quad (17)$$

   The projection vectors $\mathbf{a}_l$, $l = 1, \ldots, L$ that minimize the objective function are the eigenvectors that correspond to the L smallest eigenvalues of matrix $\sum_k \tau_k\mathbf{X}_k\left(\mathbf{L}_k + \beta\mathbf{L}_s - \gamma\mathbf{L}_d\right)\mathbf{X}_k^T - \zeta\mathbf{M}_k\mathbf{L}_{Bk}\mathbf{M}_k^T$. Finally, the projection matrix $\mathbf{A}$ is constructed: $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_L]$.
3. Next, (14), (15) are solved with respect to $\boldsymbol{\tau}$, for the projection matrix $\mathbf{A}$ that was calculated as in (17). By writing (14) in matrix form with respect to $\boldsymbol{\tau}$, we get:

$$\arg\min_{\tau} \sum_k \tau_k \left\{\mathrm{tr}\left[\mathbf{A}^T\mathbf{X}_k(\mathbf{L}_k + \beta\mathbf{L}_s - \gamma\mathbf{L}_d)\mathbf{X}_k^T\mathbf{A}\right] - \zeta \mathrm{tr}\left[\mathbf{M}_k\mathbf{L}_{Bk}\mathbf{M}_k^T\right]\right\} + \varepsilon\boldsymbol{\tau}^T\boldsymbol{\tau}, \quad (18)$$

   subject to the constraints:

$$\boldsymbol{\tau}^T\mathbf{1}_K = 1, \ \ \tau_k \geq 0, \ \ k = 1, \ldots, K, \quad (19)$$

   where $\mathbf{1}_K \in \Re^K$ is a vector of ones. The system (18)-(19) is a quadratic programming problem with respect to $\boldsymbol{\tau}$ and can be solved with any quadratic programming solver.
4. Steps 2 and 3 are repeated until convergence.

5. CLASSIFICATION VIA LABEL PROPAGATION

After the projection matrix $\mathbf{A}$ and the coefficients vector $\boldsymbol{\tau}$ are computed, the data projections $\mathbf{X}'_k$ of representation k to the reduced dimensional space are computed as:

$$\tilde{\mathbf{X}}_k = \mathbf{A}^T\mathbf{X}_k. \quad (20)$$

For each $\tilde{\mathbf{X}}_k$, a weight matrix $\tilde{\mathbf{W}}_k$ is computed according to (1), (2). Label propagation is then performed based on the weighted average:

$$\tilde{\mathbf{W}} = \sum_k \tau_k\tilde{\mathbf{W}}_k, \quad (21)$$

according to the following rule [10]:

$$l(\mathbf{x}_i) = \arg\max \mathbf{F}_{i\boldsymbol{\cdot}}, \quad (22)$$

where $\mathbf{F}_{i\boldsymbol{\cdot}}$ is the i-th raw of matrix [10]:

$$\mathbf{F} = (1-\beta)\left(\mathbf{I} - \beta\tilde{\mathbf{S}}\right)^{-1}\mathbf{Y} \in \Re^{N\times C}, \quad (23)$$

$\tilde{\mathbf{S}} = \mathrm{diag}\left(\sum_{j=1}^{M}\tilde{W}_{ij}\right)^{-1/2}\tilde{\mathbf{W}}\mathrm{diag}\left(\sum_{j=1}^{M}\tilde{W}_{ij}\right)^{-1/2}$, $0 \leq \beta \leq 1$, and $\mathbf{Y} \in \Re^{N\times C}$ is the initial state matrix, whose (i,c)-entry takes the value 1 if $l(\mathbf{x}_i) = c$, else it takes the value 0.

6. EXPERIMENTS

6.1. Stereo Facial Image Database Description

Experiments were conducted on three stereo movies. The task was to perform person identity (label) propagation on the facial images that appear in these movies with a procedure that emulates the annotation procedure followed in television archives by archivists. The movies have total duration 6 hours, 4 minutes and 16 seconds and 528,348 frames in total.

First, the movies were processed with a shot cut detection algorithm and the shot boundaries were detected. Then, the facial images where automatically extracted by performing automatic face detection and tracking. The face detector used was the Viola-Jones face detector [11], modified to incorporate color information [12] that eliminates a large amount of false detections. Face detection was performed separately on the left and right video channels, retaining only the facial images that were detected in both channels. When a facial image was detected in both channels, it was tracked for the next 20 frames or until a shot cut was detected, using a single channel appearance-based object tracker [13] resulting in a so called facial image trajectory consisting of facial image rectangular regions of interest (ROIs). The procedure was repeated for the remaining video frames. Sequential facial image trajectories that belonged to the same person were concatenated into a single trajectory. In total, 171,649 facial images were detected in 4,845 trajectories, belonging to 129 different actors plus some false detections.

Since the number of the extracted facial images is very large, the resulting graph weight matrix of the facial images would be very large and too expensive to compute. In order to decrease the computational complexity and increase annotation speed, we make the following assumptions for the data (which construct the data pairwise similarity and dissimilarity constraints):

1. facial images that belong to the same trajectory belong to the same actor,
2. facial images appearing in the same video frame belong to different actors.

According to the first assumption, only one image from each trajectory, e.g. the first, is required for the annotation process. The remaining images in the facial image trajectory simply adopt the label of the first image. However, by selecting only one image from each trajectory, information about the trajectory length is discarded during the propagation procedure. In order to retain this information, we select more images from the longer trajectories and less from the shorter ones. In total, 13,850 images were selected from the three movies, which consist 5.85% of the extracted facial images. The facial images were considered to belong to 131 classes, one class for each actor that appear in the movies. Finally, the facial images are aligned with the funnel algorithm [14].

	SSMLPP	LPP	OLPP	PCLPP	DLPP
Movie 1	**79.58%**	76.35%	72.24%	76.13%	76.90%
Movie 2	**63.17%**	58.73%	51.81%	61.70%	61.74%
Movie 3	**66.97%**	64.81%	63.77%	65.97%	66.49%
Average	**69.91%**	66.63%	62.61%	67.93%	68.38%

Table 1. Classification accuracy of SSMLPP and state the of the art LPP, OLPP, PCLPP and DLPP for three stereo videos

6.2. Comparison of SSMLPP to other subspace methods

The performance of the proposed Semi-supervised Locality Preserving Projections on multiple graphs (SSMLPP) is compared to the performance of similar state of the art subspace techniques, namely the standard Locality Preserving Projections (LPP) [1], Orthogonal Locality Preserving Projections (OLPP) [5], Locality Preserving Projections with Pairwise Constraints (PCLPP) [8] and Discriminant Locality Preserving Projections with pairwise constraints (DLPP) [7]. In the experiments, 10-fold-cross-validation was performed, were 10% of the data consist the initially labeled images. The experimental results are shown in Table 1. We notice that in all three videos the proposed SSMLPP algorithm achieves the best classification accuracy. The average increase in accuracy with SSMLPP with respect to the best state of the art subspace method, DLPP, is 1.53%.

7. CONCLUSIONS

In this paper, a novel method for propagating person identity labels on facial images extracted from stereo videos was introduced. The proposed method operates on data with multiple representations, by calculating a projection matrix that projects the multiple data representation matrices to a reduced dimensionality space that preserves the locality information in the original representations and that satisfies a priori discriminant and pairwise information in the form of pairwise must-link and cannot-link constraints. Experimental results on a large data set consisting of facial images extracted from three stereo movies showed that the subspace representation through SSMLPP has increased classification accuracy with respect to other dimensionality reduction techniques.

8. REFERENCES

[1] X Niyogi, "Locality preserving projections," in *Neural information processing systems*, 2004, vol. 16, p. 153.

[2] Limei Zhang, Lishan Qiao, and Songcan Chen, "Graph-optimized locality preserving projections," *Pattern Recognition*, vol. 43, no. 6, pp. 1993–2002, 2010.

[3] Jiwen Lu and Yap-Peng Tan, "Regularized locality preserving projections and its extensions for face recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 40, no. 3, pp. 958–963, 2010.

[4] Lishan Qiao, Songcan Chen, and Xiaoyang Tan, "Sparsity preserving projections with applications to face recognition," *Pattern Recognition*, vol. 43, no. 1, pp. 331–341, 2010.

[5] "Face recognition based on orthogonal discriminant locality preserving projections," *Neurocomputing*, vol. 70, no. 79, pp. 1543 – 1546, 2007.

[6] Weiwei Yu, Xiaolong Teng, and Chongqing Liu, "Face recognition using discriminant locality preserving projections," *Image and Vision computing*, vol. 24, no. 3, pp. 239–248, 2006.

[7] Gui-Fu Lu, Zhong Lin, and Zhong Jin, "Face recognition using discriminant locality preserving projections based on maximum margin criterion," *Pattern Recognition*, vol. 43, no. 10, pp. 3572–3579, 2010.

[8] Hakan Cevikalp, Jakob Verbeek, Frédéric Jurie, Alexander Klaser, et al., "Semi-supervised dimensionality reduction using pairwise equivalence constraints," in *3rd International Conference on Computer Vision Theory and Applications (VISAPP'08)*, 2008, pp. 489–496.

[9] Fei Wang and Changshui Zhang, "Label propagation through linear neighborhoods," in *Proceedings of the 23rd international conference on Machine learning*. 2006, ICML '06, pp. 985–992, ACM.

[10] Dengyong Zhou, Olivier Bousquet, Thomas Navin Lal, Jason Weston, and Bernhard Schlkopf, "Learning with local and global consistency," in *Advances in Neural Information Processing Systems 16*. 2004, pp. 321–328, MIT Press.

[11] Paul Viola and Michael Jones, "Robust real-time object detection," in *International Journal of Computer Vision*, 2001.

[12] GN Stamou, M Krinidis, N Nikolaidis, and I Pitas, "A monocular system for automatic face detection and tracking," in *Visual Communications and Image Processing 2005*. International Society for Optics and Photonics, 2005, pp. 59602C–59602C.

[13] S.K. Zhou, R. Chellappa, and B. Moghaddam, "Visual tracking and recognition using appearance-adaptive models in particle filters," *IEEE Transactions on Image Processing*, vol. 13, no. 11, pp. 1491–1506, 2004.

[14] Lior Wolf, Tal Hassner, and Yaniv Taigman, "Effective unconstrained face recognition by combining multiple descriptors and learned background statistics," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 10, pp. 1978–1990, 2011.