ITERATIVE LABEL PROPAGATION ON FACIAL IMAGES

Olga Zoidi, Anastasios Tefas, Nikos Nikolaidis, Ioannis Pitas

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

ABSTRACT

In this paper a novel method is introduced for propagating person identity labels on facial images in an iterative manner. The proposed method takes into account information about the data structure, obtained through clustering. This information is exploited in two ways: to regulate the similarity strength between the data and to indicate which samples should be selected for label propagation initialization. The proposed method can also find application in label propagation on multiple graphs. The performance of the proposed Iterative Label Propagation (ILP) method was evaluated on facial images extracted from stereo movies. Experimental results showed that the proposed method outperforms state of the art methods either when only one or both video channels are used for label propagation.

***Index Terms*—** label propagation, multi-graph label propagation

1. INTRODUCTION

Annotation typically aims at multimedia data archival and fast search, based on their semantic annotation (tags). This situation arises, for example, in the case of television content annotation in broadcasters' audiovisual archives. In this case, archivists usually perform a coarse annotation of the entire video, which, in many cases, is insufficient for journalists to directly access video shots/frames of interest. Such problems can be overcome with semi-automatic annotation techniques, based on label propagation [1], which is a semi-automatic process for spreading semantic labels from a small set of available labeled data to a much larger set of unlabeled data. In the case of television content annotation, one type of semantic information, which is of interest to archivists, concerns the person identities and appearances in videos to be archived. Label propagation techniques take into consideration the following assumptions: 1) visual data, e.g., video shots, frames, facial images, that are similar to each other, according to a similarity measure, or that lie in the same feature space structure (e.g., cluster, manifold) should be assigned the same label and 2) the initial labeled data should retain their label during/after label propagation.

Label propagation is performed on the visual data, according to a label inference method, which specifies the way the labels are spread from the set of labeled data to the set of unlabeled data. Usually, iterative label inference methods are employed [1]. In these algorithms, label spread is performed gradually on the unlabeled data, according to some update rule. The final label allocation converges to a stationary state, as $t \rightarrow \infty$. The stationary state of the iterative algorithm can be computed beforehand. Therefore, in such cases, these methods are performed in a single step. Such label propagation methods are introduced in [2, 3, 4]. The performance of label propagation methods depends highly on the selection of the initially labeled data set. A method for selecting the initially labeled data set is presented in [5].

There are also cases, in which multiple graphs are constructed for representing the relationships between the visual data. The fusion of multiple data representations can be performed either at the graph construction level (early fusion), e.g., by concatenating the separate feature vectors into a global feature vector, or at the decision level (late fusion), e.g., by learning a propagation algorithm for each data representation and fusing the propagation results. Late fusion is also called "multi-modal fusion" of "multi-modality learning" [6]. A study on early versus late fusion methods for semantic analysis of multi-modal video can be found in [7]. Label propagation methods on multiple graphs have been introduced in [6], [8, 9].

In this paper, we propose a novel method for label propagation on data either with a single or multiple representations that finds application in person identity label propagation on monocular/multi-view camera systems. The method first employs a subspace learning method for representing 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.

image data. If the data are obtained from multi-view camera systems, the subspace method is employed separately on all image views. Then, the data graph similarity matrix (or matrices, in the case of multi-view data) is constructed, by exploiting the information of the data structure, obtained through clustering. The clustering information is also exploited, in the initialization of the first step of the label propagation procedure. Finally, label propagation proceeds in an iterative way, by gradually adding image data in the initially labeled data set. The scope for the algorithm is to propagate facial identity information on all the facial images that appear in monocular/multi-channel videos. Experimental results showed the effectiveness of the proposed method in propagating face identity information with respect to state of the art methods.

2. ITERATIVE LABEL PROPAGATION ON SINGLE GRAPH

2.1. Cluster-based graph construction

Graph construction begins with the projection of the image data to a reduced dimensional space, maintaining the locality information of the data. More specifically, if $\mathbf{x}_i, \mathbf{x}_j \in \Re^N$ two images with strong similarity, a projection matrix $\mathbf{A} \in \Re^{N \times L}$ with $L << N$, is searched such that the data projections $\mathbf{x}'_i = \mathbf{A}^T \mathbf{x}_i$, $\mathbf{x}'_j = \mathbf{A}^T \mathbf{x}_j \in \Re^L$ are mapped close to each other.

Let $\mathbf{X}' = [\mathbf{x}'_1, \dots, \mathbf{x}'_M] \in \Re^{N \times M}$ be the data matrix and $\mathcal{G} = (\mathbf{X}', \mathcal{E})$ be the graph, whose nodes are the data matrix columns 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. This similarity is computed according to the heat kernel equation:

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

where σ is the mead edge length distance among neighbors. Information on the data structure is incorporated into the similarity graph $\mathbf{W}$, by applying a clustering algorithm, e.g., k-means clustering or n-cut [10] on the projected data. In order to increase the similarity between samples in the same cluster and suppress the similarity between samples in different clusters, the entries of the weight matrix (1) are modified as follows:

$$W_{ij} = \begin{cases} W_{ij}, & \text{if nodes } i, j \text{ belong to the same cluster} \\ \eta W_{ij}, & \text{if nodes } i, j \text{ belong to different clusters,} \end{cases} \tag{2}$$

where $0 \le \eta \le 1$ is a penalizing parameter. By setting $\eta = 0$, information transfer between different clusters is prohibited. On the other hand, if we set $\eta = 1$, no clustering information is taken into account. The influence of the parameter η selection and of the selected cluster number is examined in Section 4.1.

Instead of using the same penalizing parameter for the similarity between data in different clusters that are close to each other and clusters that are further away, we introduce a second method for re-calculating the weight matrix, which takes into account the distance between the cluster centers. More specifically, we define a new weight matrix $\mathbf{W}'$ with entries:

$$W'_{ij} = \begin{cases} W_{ij}, & \text{if nodes } i, j \text{ belong to the same cluster} \\ \zeta(c_i, c_j) W_{ij}, & \text{if nodes } i, j \text{ belong to different clusters,} \end{cases} \tag{3}$$

where W_{ij} is given by (1), c_i, c_j are the clusters of nodes i and j, respectively, and:

$$\zeta(c_i, c_j) = e^{-\frac{\|\mathbf{x}_{c_i} - \mathbf{x}_{c_j}\|^2}{\sigma}}, \tag{4}$$

where $\mathbf{x}_{c_i}$, $\mathbf{x}_{c_j}$ are the centers of clusters c_i, c_j, respectively and σ is defined as in (1). The constructed cluster-based similarity matrix is then used in label propagation.

2.2. iterative label propagation

Let us define the set of labeled data $\mathcal{X}_L = \{\mathbf{x_i}\}_{i=1}^{m_l}$, which are assigned labels from the set $\mathcal{L} = \{l_j\}_{j=1}^{Q}$ and a set of unlabeled data $\mathcal{X}_U = \{\mathbf{x_i}\}_{i=1}^{m_u}$. Without loss of generality, we define the set of labeled and unlabeled data as $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_{m_l}, \mathbf{x}_{m_l+1}, \dots, \mathbf{x}_M\}$, $M = m_l + m_u$. The vector $\mathbf{Y} = [y_1, \dots, y_{m_l}, 0, \dots, 0]^T = [\mathbf{Y}_L^T | \mathbf{Y}_U^T]^T \in \mathcal{L}^M$ contains the labels of the labeled data in the first m_l positions and takes the value 0 in the last m_u positions. The objective of label propagation methods is to spread the labels in $\mathcal{L}$ from the set of labeled data $\mathcal{X}_L$ to the set of unlabeled data $\mathcal{X}_U$. For this reason, a set of functions f_i, $i = 1, \dots, M$ is defined that assigns on the i-th graph node one value for every possible label. By defining the matrix $\mathbf{F} = [\mathbf{f}_1^T, \dots, \mathbf{f}_M^T]^T \in \Re^{M \times Q}$, label propagation is performed by the iterative process [4]:

$$\mathbf{F}^{t+1} = a\mathbf{S}\mathbf{F}^t + (1-a)\mathbf{Y}, \tag{5}$$

where:

$$\mathbf{S} = \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2} \tag{6}$$

and $\mathbf{D}$ is the diagonal matrix with $D_{ii} = \sum_j W_{ij}$. Essentially, the matrix $\mathbf{S}$ represents how much the label value of a node is affected from the label of its neighboring nodes. The parameter a, $0 \le a \le 1$, regulates the percentage of information the node will receive from its neighbors and from its initial label information. The iterative procedure (5) converges to the solution [4]:

$$\mathbf{F} = (1-a)(\mathbf{I} - a\mathbf{S})^{-1}\mathbf{Y}. \tag{7}$$

Moreover, it is proven in [4] that the iterative process given by (5) is equivalent to the manifold regularization problem:

$$\mathcal{Q}(\mathbf{F}) = \frac{1}{2}\text{tr}\left(\mathbf{F}^T\mathbf{L}\mathbf{F}\right) + \mu\text{tr}\left((\mathbf{F}-\mathbf{Y})^T(\mathbf{F}-\mathbf{Y})\right), \tag{8}$$

where $\mathbf{L} = \mathbf{D}^{-1/2}(\mathbf{D} - \mathbf{W})\mathbf{D}^{-1/2}$ is the normalized graph Laplacian and $\mu = \frac{1-a}{a}$.

The matrix $\mathbf{Y} \in \Re^{M\times Q}$ represents the initial state, with values:

$$Y_{ij} = \begin{cases} 1, & \text{if node } i \text{ is labeled as } y_i = j, \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

Label propagation performance depends highly on the initialization of matrix $\mathbf{Y}$, i.e., on the selection of the initial data samples that will be manually assigned a label. Therefore, a more structured procedure for selecting the initial labeled data set is followed, that exploits clustering information. At first, the node from each cluster with the highest within-cluster degree centrality is selected to be in the set $\mathcal{X}_L$ of initially labeled samples. The within-cluster degree centrality for a node i that belongs to cluster c is measured by summing the edge weights that connect the node i with all other nodes of cluster c:

$$d_i = \sum_{j\in\mathcal{N}_c} W_{ij}, \tag{10}$$

where $\mathcal{N}_c$ the set of graph nodes in cluster c. Intuitively, the node with the highest within-cluster degree centrality is the most representative cluster node, i.e., the node with the highest similarity to all other cluster nodes. Then, the label information is propagated to the unlabelled nodes according to the following decision rule:

$$y_i = \arg\max_j F_{ij}. \tag{11}$$

The values in $\mathbf{F}$ (11) are an indication on the "certainty" with which the node is assigned a label, i.e., nodes in which the highest F_{ij} value is much larger to the second highest F_{ij} value are more probable to be assigned the correct label, while nodes in which the two highest F_{ij} values are very close to each other, most probably lie in a 'border' region between two visual data classes. Label assignment to such nodes is more uncertain. The nodes which were assigned a label with the least certainty form the next set of nodes that will be manually labeled and inserted in the set $\mathcal{X}_L$ of labeled nodes. More specifically, for each node i we compute the difference between the two largest values in the i-th row of $\mathbf{F}$:

$$p_i = \text{order\_descend}_j(F_{ij}, 1) - \text{order\_descend}_j(F_{ij}, 2) \tag{12}$$

where $\text{order\_descend}_j(F_{ij}, k)$ is the operator that orders the rows of $\mathbf{F}$ in descending order and returns the k-th largest value. The q nodes with the smallest difference value are inserted in the set $\mathcal{X}_L$ and, the initial state matrix $\mathbf{Y}$ is updated, in order to include the newly manually labeled nodes and label propagation is performed again, according to (7) and (11). The procedure is repeated and the labeled set $\mathcal{X}_L$ is enriched with q nodes at the time with the smallest p_i value, until the cardinality of the set $\mathcal{X}_L$ is a determined percentage (e.g., 5%) of the overall data number.

3. ITERATIVE LABEL PROPAGATION ON MULTIPLE GRAPHS

3.1. cluster-based graph construction

When the image data are captured from multi-view camera systems, dimensionality reduction is performed separately in each image view, according to the chosen subspace method. Let K be the number of image views and $\mathbf{X}_k = [\mathbf{x}_{k,1}, \dots, \mathbf{x}_{k,M}] \in \Re^{N\times M}$, $k = 1\dots, K$ the data matrix of the k-th view. For each image view, a projection matrix $\mathbf{A}_k$ is computed, according to the employed subspace method and the data projections $\mathbf{X}'_k$ are computed according to $\mathbf{X}'_k = \mathbf{A}_k^T\mathbf{X}_k$. The weight matrices $\mathbf{W}_k$ are computed for each data projection k, $k = 1, \dots, K$ according to (1). The data structure of all representations is incorporated into the weight matrices by performing clustering onto the fused data representations $\tilde{\mathbf{X}}$:

$$\tilde{\mathbf{X}} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{X}'_k, \tag{13}$$

according to (2) or (3).

3.2. Iterative label propagation

After the computation of the data projections $\mathbf{X}'_k$, $k = 1, \dots, K$, label propagation is performed on the projected data by fusing the information obtained from all representations. Label propagation is performed concurrently on the K graphs, by extending the single-graph regularization framework (8) as a weighted sum of K objective functions [11]:

$$\mathcal{Q}(\mathbf{F}, \boldsymbol{\tau}) = \frac{1}{2}\sum_{k=1}^{K}\tau_k\left\{\text{tr}\left(\mathbf{F}^T\mathbf{L}_k\mathbf{F}\right) + \mu\text{tr}\left((\mathbf{F}-\mathbf{Y})^T(\mathbf{F}-\mathbf{Y})\right)\right\}, \tag{14}$$

where $\mathbf{L}_k$ is the normalized graph Laplacian of representation k. The solution of the minimization framework:

$$\arg\min_{\mathbf{F},\boldsymbol{\tau}} \mathcal{Q}(\mathbf{F}, \boldsymbol{\tau}), \quad s.t. \sum_{k=1}^{K}\tau_k = 1 \tag{15}$$

is given from the alternating solution of the equations [11]:

$$\mathbf{F} = (1-a)\left(\mathbf{I} - a\sum_k \tau_k\mathbf{S}_k\right)^{-1}\mathbf{Y}, \tag{16}$$

$$\tau_k = \frac{\left(\text{tr}(\mathbf{F}^T\mathbf{L}_k\mathbf{F}) + \mu\|\mathbf{F}-\mathbf{Y}\|^2\right)^{-1/(K-1)}}{\sum_{k=1}^{K}\left(\text{tr}(\mathbf{F}^T\mathbf{L}_k\mathbf{F}) + \mu\|\mathbf{F}-\mathbf{Y}\|^2\right)^{-1/(K-1)}}, \tag{17}$$

where $\mathbf{L}_k = \mathbf{I} - \mathbf{S}_k$, $\mathbf{S}_k = \mathbf{D}^{-1/2}\mathbf{W}_k\mathbf{D}^{-1/2}$ and $a = \frac{1}{1+\mu}$. Finally, the initialization of the labeled data set $\mathcal{X}_L$ is performed according to the iterative procedure described in Subsection 2.2.

4. EXPERIMENTS

The performance of the proposed method was tested on person identity label propagation on 13,850 stereo facial images belonging to 131 actors extracted from three stereo movies through automatic detection and tracking. A varying number of stereo facial images were extracted from each trajectory, according to the trajectory length. In total, 5,398, 3,498 and 4.954 stereo facial images were extracted from movies 1, 2 and 3, respectively. The data views are the left and right channel facial images ($K = 2$). The subspace methods used for performing dimensionality reduction are Locality Preserving Projections (LPP) [12], Orthogonal Locality Preserving Projections (OLPP) [13], Locality Preserving Projections with Pairwise Constraints (PCLPP) [14], and Neighborhood Preserving Embedding (NPE) [15]. In the case of PCLPP, the similarity and dissimilarity constraints were constructed according to the following rules:

- Two facial images are similar and, thus, are assigned the same label if they belong to the same facial image trajectory
- Two facial images are dissimilar and, thus, are assigned different labels if they are in the same frame.

The proposed method operates both on single-view and multi-view data. The performance of the proposed method on single-view data was tested on the left channel of the stereo images. The performance of the proposed method on multi-view data was tested on both channels of the stereo images.

4.1. Effect of parameters η and ζ to ILP

In this section, the effect of the prior information imposed on the projected data weight matrix obtained through clustering in the classification performance of the proposed algorithm is examined. More specifically, the classification performance was tested for parameter η values 0 (i.e., label propagation between clusters is prohibited), 0.2, 0.4, 0.6, 0.8 and 1 (i.e., no clustering information is exploited in the weight matrix). The experiment was conducted on the left image channel and the employed subspace method is LPP. The number of clusters used in the experiments was 170, 100 and 150, for the Movies 1-3. In all experiments, the data dimensionality is reduced to 75. The experimental results are shown in the first six rows of Table 1, where it is evident that the optimal classification accuracy for each movie is achieved for $\eta = 0.4$ or $\eta = 0.8$. Table 1 shows that the classification accuracy of the algorithm is sensitive to the selection of η. In all Movies, the classification accuracy is worse for $\eta = 0$, i.e., when label propagation is restricted within the clusters. Finally, we examine the classification performance when the exponential parameter ζ in (4), that takes into account the distance between the cluster centers, is exploited in label propagation. We notice that, even though the use of ζ does not lead to the best classification accuracy for all three movies, it still achieves the optimal average classification accuracy. In the rest of the experiments, clustering information will be imposed to the weight matrix with the adaptive parameter ζ.

Table 1. Significance of parameters η and ζ to the classification accuracy of iterative label propagation.

	Movie 1	**Movie 2**	**Movie 3**	**Average**
$\eta = 0.0$	71.66%	48.60%	56.59%	60.05%
$\eta = 0.2$	74.21%	53.30%	64.61%	65.53%
$\eta = 0.4$	74.60%	54.33%	**64.81%**	66.01%
$\eta = 0.6$	74.95%	55.36%	64.26%	66.20%
$\eta = 0.8$	75.18%	**56.23%**	63.89%	66.38%
$\eta = 1.0$	75.04%	56.06%	64.51%	66.56%
ζ	**75.24%**	55.98%	64.56%	**66.58%**

4.2. Effect of labeling initialization to ILP

In this section, we examine the effect of the proposed method for selecting the initially labeled data set in the classification accuracy, when dimensionality reduction is performed through various subspace methods. First, the facial images of the Movies 1-3 were divided into 170, 100 and 150 clusters, respectively and the facial images that correspond to the cluster centers were manually labeled. In each iteration of the algorithm described in Subsection 2.2, 33, 25 and 33 images in the border between clusters (that have the smallest "certainty") were assigned labels manually. The procedure was repeated 3 times. In all experiments, the initial manually labeled data set consists 5% of the facial images, i.e., for the movies 1-3 the manually labeled data set consists of 270, 175 and 250 facial images, respectively.

We compare the classification performance, when the initially labeled data set is selected randomly, without taking into account any prior information, as typically proposed in the literature. Experimental results for the case of single-view facial images are shown in Table 2 and for the case of multi-view (stereo) facial images are shown in Table 3. We notice that, in all cases, the proposed method outperforms the state of the art methods. More specifically, for the case of single-view images, the increase in classification accuracy with the proposed method is 4.71%-6.77% better than that of the state of the art. Similar results are obtained for the case of multi-view images, where the increase in achieved classification accuracy is 4.82%-6.28%.

5. 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 either on data with a single or multiple views. The method exploits information about the data structure obtained from the application of a

Table 2. Classification accuracy of the proposed single-view iterative label propagation and state of the art label propagation when LPP, OLPP, PCLPP and NPE were used for dimensionality reduction for three stereo movies

random	LPP	PCLPP	OLPP	NPE
Movie 1	71.39%	**73.21%**	66.47%	72.75%
Movie 2	53.01%	**57.17%**	46.23%	54.19%
Movie 3	59.30%	**60.25%**	57.74%	59.48%
Average	61.67%	**63.83%**	57.46%	62.54%
ILP	LPP	PCLPP	OLPP	NPE
Movie 1	75.24%	**78.19%**	72.11%	77.18%
Movie 2	55.98%	**61.78%**	51.96%	55.90%
Movie 3	64.56%	**65.39%**	64.22%	64.38%
Average	66.58%	**69.48%**	64.23%	67.25%

Table 3. Classification accuracy of the proposed multi-view iterative label propagation and state of the art label propagation when LPP, OLPP, PCLPP and NPE were used for dimensionality reduction for three stereo movies.

random	LPP	PCLPP	OLPP	NPE
Movie 1	74.26%	**75.71%**	68.40%	75.64%
Movie 2	56.05%	**59.77%**	48.68%	57.46%
Movie 3	61.50%	**62.54%**	63.06%	63.14%
Average	64.33%	**66.28%**	60.80%	65.82%
ILP	LPP	PCLPP	OLPP	NPE
Movie 1	77.77%	**80.78%**	75.02%	79.62%
Movie 2	60.18%	**65.37%**	53.33%	61.74%
Movie 3	67.59%	**68.66%**	67.32%	67.08%
Average	69.71%	**72.56%**	66.83%	70.64%

clustering algorithm. Experimental results showed that the proposed method outperforms state of the art methods either when only one or both video channels are used for label propagation.

REFERENCES

[1] X. Zhu, *Semi-Supervised Learning Literature Survey*, Technical Report, University of Wisconsin - Madison, 2008.

[2] Phi The Pham, T. Tuytelaars, and M.-F. Moens, "Naming people in news videos with label propagation," *IEEE MultiMedia*, vol. 18, no. 3, pp. 44 –55, march 2011.

[3] O Chapelle, A Zien, et al., *Semi-Supervised Learning*, MIT Press, 2006.

[4] 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.

[5] Olga Zoidi, Nikos Nikolaidis, and Ioannis Pitas, "Exploiting clustering and stereo information in label propagation of facial images," in *IEEE Symposium Series on Computational Intelligence*, 2013.

[6] Meng Wang, Xian-Sheng Hua, Richang Hong, Jinhui Tang, G.-J. Qi, and Yan Song, "Unified video annotation via multigraph learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 5, pp. 733–746, 2009.

[7] Cees G. M. Snoek, Marcel Worring, and Arnold W. M. Smeulders, "Early versus late fusion in semantic video analysis," in *Proceedings of the 13th annual ACM international conference on Multimedia*, 2005, MULTIMEDIA '05, pp. 399–402.

[8] Andreas Argyriou, Mark Herbster, and Massimiliano Pontil, "Combining graph laplacians for semi-supervised learning," in *Advances in Neural Information Processing Systems 18*. 2005, pp. 67–74, MIT Press.

[9] T. Kato, H. Kashima, and M. Sugiyama, "Robust label propagation on multiple networks," *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 35 –44, 2009.

[10] Jianbo Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888 –905, Aug 2000.

[11] Meng Wang, Xian-Sheng Hua, Richang Hong, Jinhui Tang, Guo-Jun Qi, and Yan Song, "Unified video annotation via multigraph learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 5, pp. 733–746, 2009.

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

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

[14] 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.

[15] Xiaofei He, Deng Cai, Shuicheng Yan, and Hong-Jiang Zhang, "Neighborhood preserving embedding," in *Proc. Int. Conf. Computer Vision (ICCV'05)*, 2005.