Zoidi, O., Nikolaidis, N., & Pitas, I. (2015). Label propagation on data with multiple representations through multi-graph locality preserving projections. In *2014 IEEE International Conference on Image Processing (ICIP 2014): Proceedings of a meeting held 27-30 October 2014, Paris, France* (pp. 1505-1509). (IEEE International Conference on Image Processing (ICIP)). Institute of Electrical and Electronics Engineers (IEEE).

LABEL PROPAGATION ON DATA WITH MULTIPLE REPRESENTATIONS THROUGH MULTI-GRAPH LOCALITY PRESERVING PROJECTIONS

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 propagating label information on data with multiple representations. The method performs dimensionality reduction of the data by calculating 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. The final data representations are then fused, in order to perform label propagation. The performance of the proposed method was evaluated on facial images extracted from stereo movies and on the UCF11 action recognition database. Experimental results showed that the proposed method outperforms state of the art methods.

Index Terms— Locality preserving projections, dimensionality reduction, label propagation, multiple graphs

1. INTRODUCTION

Label propagation methods aim the spread of label information from a small set of labelled data to a larger set of unlabelled data. The label propagation effectiveness depends on two factors: the graph construction and the label inference method. Graph construction deals with the choice of the data representation and the pairwise similarity (or distance) metric, while label inference methods try to assign the same label to similar data and different labels to dissimilar data, according to the above mentioned similarity measure.

A typical method for data representation mainly employed on images that leads to dimensionality reduction is Locality Preserving Projections (LPP) [1, 2]. In LPP, the data are projected to a reduced dimensionality space, so that the locality information of the original data is preserved. Several modifications of LPP exist, that incorporate additional information and constraints for the data, such as, sparsity constraints [3], discriminant information from the entire feature space [2] or from the availability of labeled data [4] and orthogonality constraints [5].

After data feature extraction through LPP and graph construction through a chosen similarity measure, label propagation is performed on the data (graph nodes), 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 [6–8]. 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 $t \to \infty$. The stationary state of the iterative algorithm can be computed beforehand. Therefore, in such cases, these methods are performed in a single step. In cases where the data can be represented in more than one feature spaces, one graph for each representation method can be constructed. The fusion of multiple data representations can be performed either at the graph construction level, e.g., by concatenating the separate feature vectors into a global feature vector, or at the decision level, e.g., by learning a propagation algorithm for each data representation and fusing the propagation results [9–11].

In this paper, we propose a novel method for label propagation on data with multiple representations. The proposed method exploits information obtained from multiple data representations, by finding a projection matrix that preserves locality information and additional a priori pairwise information between the data in all data representations in the form of must-link and cannot-link constraints. The data representations are then projected on the same reduced-dimensionality space. The data representation projections are then combined in a single optimization framework in order to perform label propagation, by extending the single-graph regularization framework in [8] as a weighted sum of multiple objective functions. The performance of the proposed method was evaluated on facial images extracted from stereo movies and on the UCF11 action recognition database.

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.

2. THE PROPOSED METHOD

We propose a novel method for performing linear dimensionality reduction on data with multiple representations, by satisfying additional pairwise similarity and dissimilarity constraints. The multiple data representations may come from the left and right channel image representations for the case of stereo visual data, or from the extraction of multiple features for representing the same data set. For each data representation, a similarity matrix is constructed. Let $\mathcal{X}_k = \{\mathbf{x}_{k,i} \in \Re^N, i = 1, \ldots, M,\ k = 1, \ldots, K\}$ be the data set in the original space and $\mathcal{G}_k = (\mathcal{X}_k, \mathcal{E}_k)$ be the graph, whose nodes are the data entries of representation k and whose edges are the pairwise data relationships. The graph weight matrix for representation k is computed according to the heat kernel equation:

$$W_{k,ij} = e^{-\frac{\|\mathbf{x}_{k,i}-\mathbf{x}_{k,j}\|^2}{\sigma_k}}, \tag{1}$$

were σ_k is the mean edge length distance among neighbors in the k-th representation. The data pairwise constraints indicate whether two data should or should not have the same label. More precisely, the sets $\mathcal{S}$ and $\mathcal{D}$ of similar and dissimilar pairs are defined as:

$$\mathcal{S} = \{(i,j)|\mathbf{x}_i, \mathbf{x}_j \text{ must have the same label}\} \tag{2}$$

and

$$\mathcal{D} = \{(i,j)|\mathbf{x}_i, \mathbf{x}_j \text{ must have different labels}\}, \tag{3}$$

respectively. The pairwise constraints are set for all data representations.

The proposed method consists of the following steps. First, the pairwise constraints are propagated to the neighboring data. Then, the projection matrix that incorporates information from all data representations and from the propagated pairwise constraints, as well as the coefficients of each representation are calculated. Finally, the data representations are projected on the new space and, subsequently, the projections are fused in an optimization framework for label propagation.

2.1. Pairwise constraints propagation

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 $\mathbf{W}_s$, $\mathbf{W}_d$, two weight matrices constructed as follows:

$$W_{s,ij} = \begin{cases} 1, & \text{if } (i,j) \in \mathcal{S} \\ 0, & \text{otherwise} \end{cases}, \quad W_{d,ij} = \begin{cases} 1, & \text{if } (i,j) \in \mathcal{D} \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

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:

$$P_{ij} = \begin{cases} \frac{1}{|\mathcal{N}_i|}, & \text{if } j \in \mathcal{N}_i \\ 0, & \text{otherwise,} \end{cases} \tag{5}$$

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:

$$\mathbf{F}_s^{(t)} = a\mathbf{P}\mathbf{F}_s^{(t-1)} + (1-a)\mathbf{W}_s, \tag{6}$$

which converges to the steady state [12]:

$$\mathbf{F}_s = (1-a)(\mathbf{I} - a\mathbf{P})^{-1}\mathbf{W}_s. \tag{7}$$

Similarly, the label dissimilarity is propagated according to:

$$\mathbf{F}_d = (1-a)(\mathbf{I} - a\mathbf{P})^{-1}\mathbf{W}_d. \tag{8}$$

The matrices $\mathbf{F}_s$ and $\mathbf{F}_d$ contain the information of the propagated pairwise constraints.

2.2. Locality Preserving Projections on Multiple Graphs

The proposed method, called Multiple-graph Locality Preserving Projections (MLPP), searches for a $N \times L$ projection matrix $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_L]$ that operates on all data representations and also searches for the optimal linear combination of the data projections. Let $\mathbf{x}_{ki}$, $k = 1, \ldots, K$ be the different data representations of sample i and $\mathbf{a}_l$, $l = 1, \ldots, L$ the projection vectors that form the columns of the projection matrix. The proposed method minimizes the objective function:

$$\begin{aligned} G(\mathbf{A}, \boldsymbol{\tau}) &= \sum_{k,l} \tau_k \Bigg\{ \sum_{i,j=1}^{M} (\mathbf{a}_l^T \mathbf{x}_{k,i} - \mathbf{a}_l^T \mathbf{x}_{k,j})^2 W_{k,ij} \\ &+ \beta \sum_{(i,j)\in\mathcal{S}} (\mathbf{a}_l^T \mathbf{x}_{k,i} - \mathbf{a}_l^T \mathbf{x}_{k,j})^2 F_{s,ij} \\ &- \gamma \sum_{(i,j)\in\mathcal{D}} (\mathbf{a}_l^T \mathbf{x}_{k,i} - \mathbf{a}_l^T \mathbf{x}_{k,j})^2 F_{d,ij} \Bigg\} \\ &+ \varepsilon \|\boldsymbol{\tau}\|^2, \end{aligned} \tag{9}$$

subject to the constraints:

$$\mathbf{a}_l^T \mathbf{a}_j = \delta_{lj}, \ \sum_k \tau_k = 1, \ \tau_k \geq 0, \tag{10}$$

where τ_k, $k = 1, \ldots, K$ is the weight of the k-th data representation in the optimization framework, β, γ are parameters that regulate the significance of the pairwise similarity and dissimilarity constraints, respectively and ε is a regularization parameter that prevents the coefficients vector $\boldsymbol{\tau}$ from taking increased value for only one image representation. The first sum in (9) ensures that the locality information of the data in the original space is preserved in the projected space. The second/third sum in (9) ensure that the similar/dissimilar data pairs are projected close to/away from each other. Finally, the first constraint in (10) ensures that the projection matrix $\mathbf{A}$ is orthonormal. By simple algebraic manipulations, (9) can be written as:

$$\arg\min_{\mathbf{a}_l,\boldsymbol{\tau}} \sum_{k,l} \tau_k \mathbf{a}_l^T \mathbf{X}_k \left(\mathbf{L}_k + \beta \mathbf{L}_s - \gamma \mathbf{L}_d\right) \mathbf{X}_k^T \mathbf{a}_l + \varepsilon \|\boldsymbol{\tau}\|^2, \tag{11}$$

where $\mathbf{L}_k = \mathbf{D}_k - \mathbf{W}_k$ is the graph Laplacian for the k-th data representation and $\mathbf{L}_s = \mathbf{D}_s - \mathbf{F}_s$, $\mathbf{L}_d = \mathbf{D}_d - \mathbf{F}_d$ are the graph Laplacians of the pairwise similarity and dissimilarity constraints, respectively. $\mathbf{L}_k$ varies according to the data representation, while $\mathbf{L}_s$, $\mathbf{L}_d$ are constant for all representations. By selecting the parameters β, γ so that the matrix $\mathbf{L}_c = \beta\mathbf{L}_s - \gamma\mathbf{L}_d$ is positive semi-definite, the cost function (11) under the constraints (10) is convex, with respect to the variables $\mathbf{a}_l$ and $\boldsymbol{\tau}$. Then, the optimization problem is solved iteratively for $\mathbf{a}_l$ 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 (10), (11) is solved for $\mathbf{a}$ by constructing the Lagrangian function:

$$\mathcal{L}(\mathbf{a}_l, \lambda) = \mathbf{a}_l^T \left[\sum_k \tau_k \mathbf{X}_k \left(\mathbf{L}_k + \beta \mathbf{L}_s - \gamma \mathbf{L}_d\right) \mathbf{X}_k^T\right] \mathbf{a}_l - \lambda \mathbf{a}_l^T \mathbf{a}_l. \tag{12}$$

   By setting the partial derivative of the Lagrangian function with respect to $\mathbf{a}_l$ equal to zero $\frac{\partial \mathcal{L}(\mathbf{a}_l,\lambda)}{\partial \mathbf{a}_l} = 0$, we get:

$$\left[\sum_k \tau_k \mathbf{X}_k \left(\mathbf{L}_k + \beta \mathbf{L}_s - \gamma \mathbf{L}_d\right) \mathbf{X}_k^T\right] \mathbf{a}_l = \lambda \mathbf{a}_l. \tag{13}$$

   It is easy to see that 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 $\sum_k \tau_k \mathbf{X}_k \left(\mathbf{L}_k + \beta \mathbf{L}_s - \gamma \mathbf{L}_d\right) \mathbf{X}_k^T$. Finally, the projection matrix $\mathbf{A}$ is constructed: $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_L]$.

3. Next, (10), (11) are solved with respect to $\boldsymbol{\tau}$, for the projection matrix $\mathbf{A}$ that was calculated as in (13). By writing (11) in matrix form with respect to $\boldsymbol{\tau}$, we get:

$$\arg\min_{\tau} \sum_k \tau_k \mathrm{tr}\left[\mathbf{A}^T \mathbf{X}_k (\mathbf{L}_k + \beta \mathbf{L}_s - \gamma \mathbf{L}_d) \mathbf{X}_k^T \mathbf{A}\right] + \varepsilon \boldsymbol{\tau}^T \boldsymbol{\tau}, \tag{14}$$

   subject to the constraints:

$$\boldsymbol{\tau}^T \mathbf{1}_K = 1, \;\; \tau_k \geq 0, \;\; k = 1, \ldots, K, \tag{15}$$

   where $\mathbf{1}_K \in \Re^K$ is a vector of ones. The system (14)-(15) 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.

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:

$$\mathbf{X}'_k = \mathbf{A}^T \mathbf{X}_k. \tag{16}$$

2.3. Label propagation on multiple graphs

The data projections $\mathbf{X}'_k$ are fused, in order to perform label propagation. First, one graph is constructed for each $\mathbf{X}'_k$, with weight matrices $\mathbf{W}'_k$ computed from (1). Then, label propagation is performed concurrently on the K graphs, by extending the single-graph regularization framework in [8] as a weighted sum of K objective functions:

$$\mathcal{Q}(\mathbf{F}) = \frac{1}{2}\sum_{k=1}^{K} \tau_k \mathrm{tr}\left(\mathbf{F}^T \mathbf{L}'_k \mathbf{F}\right) + \frac{1}{2}\mu \mathrm{tr}\left((\mathbf{F} - \mathbf{Y})^T (\mathbf{F} - \mathbf{Y})\right), \tag{17}$$

where $\mathbf{L}'_k$ is the normalized graph Laplacian of representation k. The weights τ_k are determined as in Section 2.2. (17) is convex with respect to $\mathbf{F}$. Therefore, the global optimum can be found by setting the partial derivative of $\mathcal{Q}(\mathbf{F})$ equal to zero. The global optimum is thus given by:

$$\mathbf{F} = (1 - \zeta)\left(\mathbf{I} - \zeta \sum_k \tau_k \mathbf{S}'_k\right)^{-1} \mathbf{Y}, \tag{18}$$

where we set $\mathbf{L}'_k = \mathbf{I} - \mathbf{S}'_k$, $\mathbf{S}'_k = \mathbf{D}'^{-1/2}_k \mathbf{W}'_k \mathbf{D}'^{-1/2}_k$ and $\zeta = \frac{1}{1+\mu}$.

3. EXPERIMENTS

3.1. Label propagation on stereo facial images

The performance of the proposed method was tested on person identity label propagation on 13,850 facial images belonging to 131 actors extracted from three stereo movies through automatic detection and tracking which can be performed in various ways [13–15]. More specifically, 5,398, 3,498 and 4.954 facial images were extracted from movies 1, 2 and 3, respectively. The data modalities are the left and right channel facial images ($K = 2$). The similarity and dissimilarity weight matrices (4) are constructed as:

$$W_{s,ij} = \begin{cases} 1, & \text{if images } i,j \text{ are in the same trajectory} \\ 0, & \text{otherwise,} \end{cases} \tag{19}$$

$$W_{d,ij} = \begin{cases} 1, & \text{if images } i,j \text{ are in the same frame} \\ 0, & \text{otherwise.} \end{cases} \tag{20}$$

Table 1. Classification accuracy of MLPP and single-channel LPP, OLPP, PCLPP and NPE for three stereo movies

	MLPP	LPP	OLPP	PCLPP	NPE
Movie 1	**83.47**%	76.17%	71.32%	78.34%	77.23%
Movie 2	**67.43**%	59.29%	51.95%	62.21%	60.18%
Movie 3	**71.21**%	64.91%	63.69%	66.01%	65.60%
Average	**75.05**%	67.90%	63.73%	69.03%	68.78%

Table 2. Classification accuracy of MLPP and stereo LPP, OLPP, PCLPP and NPE for three stereo movies

	MLPP	LPP	OLPP	PCLPP	NPE
Movie 1	**83.47**%	78.40%	73.11%	80.36%	79.48%
Movie 2	**67.43**%	62.40%	54.68%	64.36%	63.62%
Movie 3	**71.21**%	67.65%	66.78%	68.54%	68.30%
Average	**75.05**%	70.53%	66.22%	72.10%	71.49%

The performance of the proposed Locality Preserving Projections on multiple graphs (MLPP) 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) [16] and Neighborhood Preserving Embedding (NPE) [17]. In the experiments, 10-fold cross validation was employed. One fold was manually assigned labels and the labels were propagated to the rest. The dimension of the data is reduced from 1200 to 75. In order to test the significance of the stereo information to the classification accuracy, we compared the performance of the proposed algorithm to the performance of LPP, OLPP, PCLPP and NPE when they operate on one luminance channel of the stereo video. Label propagation is performed by exploiting local and global consistency, as proposed in [8]. The experimental results are shown in Table 1. We notice that in all three videos the classification accuracy of the proposed MLPP algorithm achieves a much better classification accuracy. The average increase in accuracy with MLPP with respect to the best single-channel subspace method PCLPP is 6.02%.

Next, we test the performance of the single-channel subspace methods, when they operate separately on the left and right channels of the stereo videos and the late fusion method described in [18] is employed for performing label propagation on the stereo facial images. The experimental results are shown in Table 2. We notice that, when the existing dimensionality reduction techniques are combined with the label propagation approach and a late fusion approach the performance is again worse than the performance of the proposed MLPP algorithm. More specifically, the average increase in accuracy with MLPP with respect to the best state of the art stereo method is 2.95%.

3.2. Algorithm performance on data with more modalities

The proposed method has been tested in the UCF11 data set [19], that consists of 1,600 Youtube videos depicting 11 action classes. Each video is represented with the state of the art action description exploiting the BoF-based video representation [20] evaluated on 5 descriptor types, each description type consisting one data modality ($K = 5$): Histograms of Oriented Gradients (Mod 1), Histograms of Optical Flow (Mod 2), Motion Boundary Histograms projected on the x- (Mod 3) and y-axis (Mod 4) and Normalized Trajectories (Mod 5). In the experiments, 10-fold cross validation was employed. One fold was manually assigned labels and the labels were propagated to the rest. The dimension of the data is reduced from 1000 to 75. Since no prior pairwise constraints information is available for the data, the matrices $\mathbf{L}_s$, $\mathbf{L}_d$ are set zero. The experimental results for the proposed method and the state of the art method LPP, which achieved the best performance, are shown in Table 3. The experimental results for the single-modality methods are shown in Table 4. We notice that, the performance of MLPP is 6.95% better than the performance of the best single-modality LPP and 0.3% better than the multi-channel LPP.

Table 3. Classification accuracy of MLPP and multi-modal LPP, OLPP and NPE.

MLPP	LPP	OLPP	NPE
64.92%	64.62%	59.24%	59.99%

Table 4. Classification accuracy of LPP for modalities 1-5.

	Mod 1	Mod 2	Mod 3	Mod 4	Mod 5
LPP	56.87%	48.74%	57.97%	57.59%	48.76%
OLPP	48.06%	47.01%	42.63%	44.25%	39.47%
NPE	33.47%	49.00%	38.36%	56.34%	48.33%

4. CONCLUSION

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 pairwise information in the form of pairwise must-link and cannot-link constraints. Experimental results showed that the proposed MLPP has increased classification accuracy compared to state of the art methods.

5. REFERENCES

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

[2] J. Lu and Y.-P. 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.

[3] L. Qiao, S. Chen, and X. Tan, "Sparsity preserving projections with applications to face recognition," *Pattern Recognition*, vol. 43, no. 1, pp. 331–341, 2010.

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

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

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

[7] P. T. 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.

[8] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schlkopf, "Learning with local and global consistency," in *Advances in Neural Information Processing Systems 16*. MIT Press, 2004, pp. 321–328.

[9] M. Wang, X.-S. Hua, R. Hong, J. Tang, G.-J. Qi, and Y. Song, "Unified video annotation via multigraph learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 5, pp. 733–746, 2009.

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

[11] 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.

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

[13] O. Zoidi, A. Tefas, and I. Pitas, "Visual object tracking based on local steering kernels and color histograms," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 5, pp. 870–882, May 2013.

[14] M. Pateraki, H. Baltzakis, P. Kondaxakis, and P. Trahanias, "Tracking of facial features to support human-robot interaction," in *IEEE International Conference on Robotics and Automation, 2009. ICRA '09*, May 2009, pp. 3755–3760.

[15] I. Pitas, *Digital video and television*. I. Pitas, 2013.

[16] H. Cevikalp, J. Verbeek, F. Jurie, A. 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.

[17] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," in *Proc. Int. Conf. Computer Vision (ICCV'05)*, 2005.

[18] M. Wang, X.-S. Hua, R. Hong, J. Tang, G.-J. Qi, and Y. Song, "Unified video annotation via multigraph learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 5, pp. 733–746, 2009.

[19] J. Liu, J. Luo, and M. Shah, "Recognizing realistic actions from videos "in the wild"," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2009, pp. 1996–2003.

[20] H. Wang, A. Kläser, C. Schmid, and C.-L. Liu, "Dense trajectories and motion boundary descriptors for action recognition," *International Journal of Computer Vision*, vol. 103, no. 1, pp. 60–79, 2013.