Fast Constrained Person Identity Label Propagation in Stereo Videos Using a Pruned Similarity Matrix

Efstratios Kakaletsis, Olga Zoidi, Ioannis Tsingalis, Anastasios Tefas, Nikos Nikolaidis, Ioannis Pitas

Artificial Intelligence & Information Analysis Lab
Department of Informatics
Aristotle University of Thessaloniki
Box 451, GR-54124 Thessaloniki, GREECE

Abstract

In this paper, a novel video data (more specifically facial images) fast labeling method, that aims in the acceleration of a state of the art facial identity label propagation technique is presented. Our method assumes that facial images are derived by applying facial image tracking on stereoscopic videos and thus are temporally ordered. The proposed method utilizes a pruned similarity matrix so that the facial label inference is conducted using fewer entries in this matrix, namely the pairwise similarities of the facial images that exist in the main and the N upper and lower off-diagonals. The proposed method can also incorporate pairwise facial image similarity and dissimilarity constraints into the objective function of the label propagation. Experiments which have been conducted on facial image labeling in three stereoscopic movies, confirm the increased labeling accuracy and the reduced computational cost of the proposed method.

Keywords:
label propagation, pruned similarity matrix, pairwise constraints

Email address: {tefas, nikolaid, pitas}@aiia.csd.auth.gr (Efstratios Kakaletsis, Olga Zoidi, Ioannis Tsingalis, Anastasios Tefas, Nikos Nikolaidis, Ioannis Pitas)

1. Introduction

The volume of images and videos captured, stored, transmitted and shared (through e.g. Facebook, Instagram and YouTube) increases with tremendous rates nowadays. Semantic annotation (tagging) of such data with respect to objects, persons, actions, etc depicted in them is thus becoming a real necessity. Such an annotation can facilitate their search, organization, retrieval and browsing. Obviously, annotation of videos is far more challenging and cumbersome compared to the annotation of images, due to the larger volume of information contained in them. Fully manual per-frame annotation/tagging of videos is essentially impractical even for small collections. Supervised or semi-supervised learning /classification approaches can greatly facilitate the video annotation task and a great progress has been achieved in both areas during the recent past. Deep learning has lately dominating the interest of researchers in this area due to the superior performance achieved by the related algorithms [1, 2, 3]. However, end-to-end deep learning approaches (i.e., approaches that perform both feature extraction and classification in a deep learning fashion) require a very large number of training examples which might not be available in certain applications. In such cases, "shallow" learning/classification methodologies are perhaps the only feasible approaches. Label propagation techniques, which spread labels from a small labeled dataset (e.g. a dataset of facial images) to a large unlabeled one, based on their similarity according to some set of selected features, is a promising and widely used approach that falls within the general category of semi-supervised learning. It should be noted however that label propagation can be combined with deep learning in the sense that the features involved in similarity calculation could very well be deep features evaluated from a pre-trained model such as those in [4] in the case of face recognition task.

A fast person identity label propagation method applicable to facial images derived from stereoscopic videos is proposed in this paper. The method incorporates a pruned similarity matrix so as to reduce computational complexity as well as pairwise constraints, towards increasing the labeling accuracy.

A multitude of label propagation approaches have been proposed in the literature. A detailed but rather old survey on semi-supervised learning can be found in [5]. Label propagation can be formally described using graphs whose nodes and edges represent the data (e.g. facial images) representations and their pairwise similarities, respectively [6]. Then, the label infer-

ence is an information diffusion process from labeled nodes to unlabeled ones through the graph paths. A recent review of graph-based label propagation approaches in digital media can be found in [7]. In iterative label propagation methods, label diffusion is performed gradually on the unlabeled data, according to an update rule [5]. A method for label propagation on similarity graphs can be found in [8]. The proposed algorithm, called Linear Neighborhood Propagation (LNP), can discover the structure of the whole dataset through synthesizing the linear neighborhood around each data object. Moreover, in [9] the learning problem is formulated through a Gaussian random field on the graph, where the mean of the field model is characterized in terms of harmonic functions, and is efficiently obtained using matrix methods or belief propagation. Another popular label propagation method is the local and global consistency method [10] which is a principled approach to semi-supervised learning that presents a classification function that is sufficiently smooth with respect to the intrinsic structure collectively revealed by known labeled and unlabeled points. Its Multiple-graph Locality Preserving Projections - Cluster-based Label Propagation (MLPP-CLP) variant is described in [11].

In [12] the authors discuss the problem of robust visual representation and label prediction proposing a new Discriminative Sparse Flexible Manifold Embedding model with a novel graph weight construction method by integrating class information and considering a certain kind of similarity/dissimilarity of samples so that the true neighborhoods can be discovered. In [13] a novel adaptive label propagation approach with joint discriminative clustering on manifolds for representing and classifying high-dimensional data is proposed. The method is capable of propagating label information using adaptive weights over low-dimensional manifold features incorporating the adaptive graph weight construction with label propagation. Essentially, the method combines unsupervised manifold learning, discriminative clustering and adaptive classification into a unified model. In [14] the authors propose a technique to encode the neighborhood reconstruction error more accurately and reliably using the nuclear norm that has been proved to be more robust to noise and more suitable to model the reconstruction error in label prediction. In [15] the authors propose the Transductive Classification Robust Linear Neighborhood Propagation (R-LNP) method that encodes the neighborhood reconstruction error more accurately by applying the L2,1-norm. The authors also use a weighted L2,1-norm regularization on the fitness error to achieve robustness to noise and more discriminatory power.

[16] proposes a new transductive label propagation method, called Adaptive Neighborhood Propagation that combines sparse coding and neighborhood propagation into a single framework. A procedure to construct the so-called directed l_1-graph, in which vertices involve all the samples and the ingoing edge weights to each vertex describe its l_1-norm driven reconstruction from the other samples and the noise is proposed in [17]. The graph can be used in various machine learning tasks, including label propagation. In [18] four strategies for graph construction that are based on adaptive sparse sample reconstruction and use the data self-representativeness property are proposed. The methods avoid l_1 coding and rely on collaborative representation adopting locality-constrained linear coding (LLC) [19]. The [20] introduces a data-driven graph construction method that exploits and extends the Local Hybrid Coding (LHC) scheme [21] which blends sparsity and bases-locality criteria in a unified optimization problem and can retain the strengths of both sparsity and locality. The proposed approach has been successfully used in the task of graph-based label propagation. In [22] a framework that can make any graph construction method incremental, i.e. add new samples (labeled or unlabeled) to a previously constructed graph, is proposed. As a case study the authors apply their algorithm to the Two Phase Weighted Regularized Least Square (TPWRLS) graph construction method [23]. In [24] the authors use as a starting point the Adsorption algorithm [25] and modify it, while reatining its desirable properties, to come up with the Modified Adsorption (MAD) method. The learning problem is formulated as an optimization problem and efficient iterative methods are developed to solve it. MAD can also be easily extended to handle data with non-mutually exclusive labels. The recently proposed OMNI-Prop method [26] performs label propagation on the graph by assigning each node with the prior belief and subsequently updating it by using the evidence from its neighbors and considering the shape of the probability distribution. The method can tackle arbitrary label correlations (e.g. homophily and heterophily) and was shown to achive better performance than classic label propagation approaches. Moreover, in [27], the authors are mainly focusing on how to incorporate the classification confidence in order to enhance accuracy and how to handle both homophily and heterophily networks in an unified framework with data characterized by arbitrary label correlations. The proposed algorithm is called Condence-Aware Modulated Label Propagation (CAMLP). Finally, the Multi-view Learning with Adaptive Neighbours (MLAN) method proposed in [28] comprises a novel multi-view learning model which performs semi-supervised classification and

local structure learning simultaneously and can allocate ideal weights for each view automatically without additional weight and penalty parameters, while avoiding to include noise and outlying entries that result in unreliable and inaccurate graphs.

The main aim of the proposed video label propagation method is the speedup of the state of the art MLPP-CLP label propagation method [11] through an approximate label propagation approach using a pruned facial image similarity matrix. A number of approaches have been proposed for decreasing the computational complexity of label propagation. Indeed, since the number of the facial images is sometimes very large, the resulting facial image similarity matrix might be too large and expensive to compute. Thus, approximate methods reduce the similarity matrices dimensions,[29, 30, 31]. Another approach is to remove the weakest facial image similarity entries from the similarity matrix while retaining only similarities within a facial image neighborhood [32, 33, 34].

The proposed label propagation technique reduces the computational complexity by utilizing a sub-sampled (pruned) similarity matrix resulting from the respective pruned similarity graph. More specifically, instead of the full facial image similarity matrix, only its main diagonal and some off-diagonal entries, are used, by exploiting the temporal ordering of facial images, that results from the fact that they are derived from temporally ordered video frames. Indeed, the facial images used as input in the algorithm are extracted by performing face detection and tracking in the two views of a stereo video [11], resulting in the so called facial image trajectories. The facial image trajectories consist of the temporally ordered regions of interest (ROIs) representing detected rectangular facial images/regions of size $N_x \times N_y$ pixels. Thus, the fact that only image similarities residing in a band around the main diagonal of the similarity matrix are calculated, implies that we take into account only similarities between temporally nearest neighbors. The rationale behind considering only similarities between facial images that are temporally close within a video (and assuming that similarities between temporally distant images are zero) lies in the way people usually appear in videos, in particular actors in a movie. Since a movie is temporally organized in scenes, each of which presents some sort of action taking place in a single location and being continuous in time, it is highly possible to have actors appearing in multiple, closely spaced in time, instances. Such an example is a dialogue scene where the camera alternates between the two dialogue participants. Thus by labeling a few instances of an actor and keeping only the

similarities between temporally close facial images one can expect to be able to successfully propagate labels. Even if we ignore (set to zero) similarities between facial images being far apart in terms of time, we can still expect that the label information will find paths through the graph to reach distant samples. This is particularly true for actors in leading roles which appear frequently and usually in a uniformly distributed (over time) way throughout a movie.

The second aim of the paper is enhancing labeling accuracy, by incorporating pairwise facial image similarity and dissimilarity constraints into the objective function of the label propagation. The idea of incorporating pairwise constraints into graph label propagation is also addressed in a number of papers. [35] presents a graph-based learning approach to pairwise constraint propagation on multi-view data. The authors decompose the inter-view (across views) constraint propagation problem into semi-supervised learning subproblems so that they can be solved using graph-based label propagation. The [36] proposes a novel way to generate constraints from the propagated labels in constrained clustering style algorithms or in label propagation algorithms. In [37] a new graph based constrained semi-supervised learning (G-CSSL) framework is proposed, using pairwise constraints to specify the types (intra- or inter-class) of points with labels. The core idea is to create and enrich the pairwise constraints sets using the propagated soft labels from the labeled and unlabeled data by special label propagation, thus obtaining more supervised information.

The proposed approach utilizes positive (similarity) constraints of the form 'facial images i, j belong to the same actor' and negative (dissimilarity) constraints of the form 'facial images i j do not belong to the same actor' [11]. The incorporation of such pairwise similarity and dissimilarity constraints into the objective function of the label propagation leads to an increased face recognition accuracy as proven in the experiments. Note that MLPP-CLP already incorporates such constraints. However, MLPP-CLP uses these constraints in the Locality Preserving Projections that it incorporates in order to achieve dimensionality reduction, whereas in our case these constraints are additionally incorporated in the label propagation objective function. This is another important contribution of this paper.

Essentially, the paper proposes three different approaches:

- a fast but still well performing (in terms of classification accuracy) version of MLPP-CLP that involves a pruned similarity matrix (PB-

MLPP-CLP)

- an improved, in terms of classification accuracy, variant of MLPP-CLP that incorporates similarity and dissimilarity constraints in the objective function (CMLPP-CLP) and

- a combination of the above schemes i.e. a fast and , in most cases, better performing version of MLPP-CLP that incorporates a pruned similarity matrix and pairwise constraints in the objective function (PCB-MLPP-CLP).

It should be noted here that the proposed methods can be used to perform label propagation not only on facial images but also on other types of image data derived from videos, such as images of objects. However, these images should be temporally ordered, for example result from a video tracking procedure which will induce temporal ordering, since the utilized similarity matrix pruning approach is based on such an ordering.

The rest of this paper is organized as follows: Section 2 provides an overview on the state of the art MLPP-CLP label propagation method [11]. Section 3 describes the details of the proposed method that implements the pruned label propagation on facial images incorporating the pairwise similarity and dissimilarity constraints. In Section 4, we present the facial images datasets and the experiments which have been conducted to measure the facial recognition accuracy and the reduced computational complexity. Finally, conclusions are presented in Section 5.

2. MLPP-CLP facial image label propagation

A very short description of the MLPP-CLP approach is presented in this section. The full algorithm can be found in [11],[6].

Assume a set of labeled facial images $X_L = \{\mathbf{x}_i\}_{i=1}^{m_l}$ which have been assigned labels (actor names) from the set $L = \{l_j\}_{j=1}^{Q}$ and a set of unlabeled facial images $X_U = \{\mathbf{x}_i\}_{i=1}^{m_u}$. Their union is given by $X = \{\mathbf{x}_1, ..., \mathbf{x}_{m_l}, \mathbf{x}_{m_l+1}, ..., \mathbf{x}_M\}, M = m_l + m_u$ [7]. The objective of label propagation is to spread the facial image labels in L from the set of the labeled images X_L to the set of the unlabeled images X_U, while maintaining local and global labeling consistency [10]. The initial information about the labeled data is described by the $M \times Q$ matrix

$\mathbf{Y}$, defined as:

$$Y_{ij} = \begin{cases} 1, & \text{if image } i \text{ is labeled as } y_i = j \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

The algorithm begins with the construction of a symmetric facial image similarity matrix $\mathbf{W} \in \mathbb{R}^{M \times M}$, as described in [6], which represents the k-nearest neighbor (k-NN) facial image similarity graph. This matrix will be called from now onwards full similarity matrix. In our case, the rows/columns of the matrix correspond to the temporally ordered facial images, i.e. the facial images in the sequence they appear in the video. The element W_{ij} of this matrix denotes the similarity between the i-th and the j-th facial image. More specifically, the edge in the graph that connects the nodes (facial images) i and j is assigned with a value W_{ij} that indicates the similarity between these two nodes. This similarity is computed according to the heat kernel equation:

$$W_{ij} = \begin{cases} e^{-\frac{\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2}{2\sigma^2}}, & i \neq j, \mathbf{x}_i, \mathbf{x}_j \text{ are k-NN} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where σ is the mean distance among neighbors and x_i is the feature vector used to represent node/image i. The main diagonal $W_{ii}, i = 1, \ldots, M$ of the similarity matrix consists of the similarities of facial images with themselves which is set equal to zero because there is no point to conduct label propagation from a facial image to itself. The construction of such a matrix has computational complexity and memory requirements of the order $O(M^2)$ even if a k nearest neighbor matrix [11] is computed.

Then, vectors $\mathbf{f}_i \in \mathbb{R}^{1 \times Q}, i = 1, ..., M$ that assign a score for every possible actor label to facial image i, thus defining the matrix $\mathbf{F} = [\mathbf{f}_1^T, ..., \mathbf{f}_M^T]^T \in \mathbb{R}^{M \times Q}$, are calculated. More specifically, $\mathbf{F}$ is calculated by minimizing [6]:

$$Q(\mathbf{F}) = \frac{1}{2} tr(\mathbf{F}^T \mathbf{L} \mathbf{F}) + \mu tr((\mathbf{F} - \mathbf{Y})^T (\mathbf{F} - \mathbf{Y})), \quad (3)$$

where $\mathbf{L} = \mathbf{D}^{-1/2}(\mathbf{D} - \mathbf{W})\mathbf{D}^{-1/2}$ is the normalized facial image similarity graph Laplacian, $\mathbf{D}$ is the diagonal matrix having entries $D_{ii} = \sum_j W_{ij}$ and μ is a regularization parameter. This minimization problem leads to the following solution:

$$\mathbf{F} = (1 - a)(\mathbf{I} - a\mathbf{S})^{-1}\mathbf{Y}, \quad (4)$$

where $a = \frac{1}{1+\mu}$ and:

$$\mathbf{S} = \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}, \tag{5}$$

The final facial image label (actor name) is assigned to facial image i according to the following decision rule:

$$y_i = \arg\max_{j \in 1,...,Q} [f_{i1}, \ldots, f_{ij}, \ldots, f_{iQ}]. \tag{6}$$

The regularization framework (3) can be easily extended to the case of label propagation on multiview facial images. In this case, multiple graphs are constructed for the data, one for each one of the K facial image representations (i.e., views, for example two views, $K = 2$, for stereoscopic images). Each graph is represented by the corresponding similarity matrix $\mathbf{W}_k, k = 1, ..., K$. In this case, the regularization framework (3) takes the form:

$$Q(\mathbf{F}, \boldsymbol{\tau}) = \frac{1}{2} \sum_{k=1}^{K} \tau_k tr(\mathbf{F}^T \mathbf{L}_k \mathbf{F}) + \mu tr((\mathbf{F} - \mathbf{Y})^T (\mathbf{F} - \mathbf{Y})), \tag{7}$$

subject to the constraint:

$$\sum_{k=1}^{K} \tau_k = 1, \tag{8}$$

that leads to the following optimal solution for $\mathbf{F}$:

$$\mathbf{F} = (1 - a) \left(\mathbf{I} - a \sum_k \tau_k \mathbf{S}_k \right)^{-1} \mathbf{Y}. \tag{9}$$

where τ_k, $k = 1,...,K$ is the weight that corresponds to the k-th data representation and $\mathbf{S}_k = \mathbf{D}^{-1/2}\mathbf{W}_k\mathbf{D}^{-1/2}$.

A method for computing the weights τ_k called Multi-graph Locality Preserving Projections (MLPP) was introduced in [11], being a variant of the Locality Preserving Projections (LPP) method [38]. It performs dimensionality reduction [39] of data with multiple representations by constructing a single projection matrix $\mathbf{A}$ for all data representations, while preserving the data locality information in all representations and ensuring additional pairwise similarity and dissimilarity constraints on the data [40]. The weights τ_k of each data representation to the construction of the projection matrix $\mathbf{A}$ are the optimal weights for the label propagation cost function (7), given

that the data feature extraction was performed according to MLPP. Once MLPP projection is performed, normalized graph Laplacians $\mathbf{L}_k$ in (7) refer to the projected data. More details about the method can be found in [11].

3. Pruned Constrained Label Propagation

3.1. Pruned Label Propagation

The proposed novel label propagation facial image technique employs a pruned facial image similarity matrix $\mathbf{W}$, instead of a full (k-NN) matrix. Despite the fact that the proposed technique is a well known method in mathematics (utilizing the band matrix for accelerating the solution of a linear system), the temporal order of the facial images in the facial image trajectories which are derived from face detection and especially from face tracking in consecutive frames, is exploited here. More specifically, the rows/columns of the matrix correspond to the temporally ordered facial images, i.e. the facial images in the sequence they appear in the video, as we had already mentioned. We assume that all images in a facial image trajectory correspond to the same person and thus for the label propagation we use only a few images of each trajectory (more details on image selection method used are provided in Section 4). Using more than one images per trajectory, instead of one, allows to take into account the temporal variations of facial appearances. The similarities of these images (in the main diagonal) and the temporally nearest neighbours (in off-diagonals) are stored in the band of the similarity matrix. The remaining images in each facial image trajectory adopt the label assigned to the obtained few images of the trajectory by the label propagation procedure. Thus, the proposed method, called Pruned Band MLPP-CLP (PB-MLPP-CLP), is accomplished using the following approach:

3.1.1. Band similarity matrix

The main diagonal $W_{ii}, i = 1, \ldots, M$ of the similarity matrix consists of the similarities of facial images with themselves and is set zero because there is no point to conduct label propagation from a facial image to itself. Around these diagonal elements, we calculate only the entries of the N upper and lower diagonals that contain the similarities of the temporally nearest neighbouring facial images as it is shown in the Figure 1. The similarity

matrix is computed according to the Gaussian heat kernel equation:

$$W_{ij} = \begin{cases} e^{-\frac{\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2}{2\sigma^2}}, & i \neq j, \mathbf{x}_i, \mathbf{x}_j \text{ are k-NN and} \\ & \in N \text{ upper/lower diagonals} \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

where $\mathbf{x}_i, \mathbf{x}_j \in \Re^M$ are the feature vectors of the i-th and j-th facial images and σ is a diffusion parameter. Obviously, $W_{ij} = W_{ji}$. Furthermore, for a band similarity matrix of the form (10), $\mathbf{S}$ (5) is a band matrix as well. In case of data with multiple representations (i.e. stereo data) where multiple similarity matrices exist (see previous section) the above procedure is applied in each matrix separately.

The experiments have shown that the classification accuracy using either the full (namely the k-NN similarity matrix) or the band similarity matrix for label propagation is approximately the same. Figure 2, illustrates a full (namely 10-NN) and a band similarity matrix.

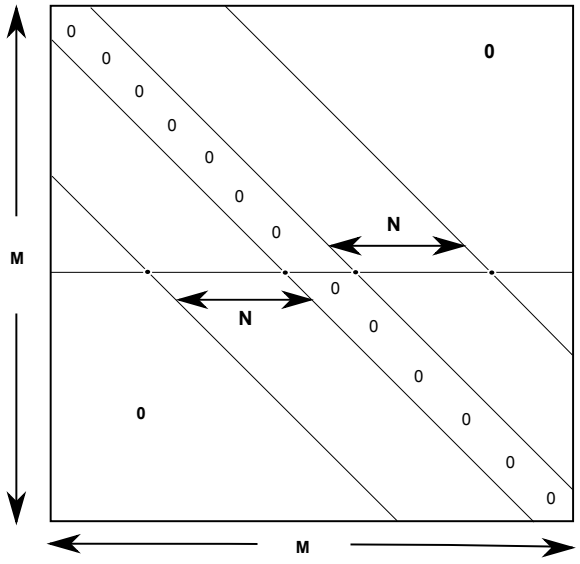

Figure 1: $(2N + 1)$-band Similarity matrix

3.2. Computational Complexity Study

The construction of a band similarity matrix has computational complexity $O(2NM) \simeq O(NM)$, which is much less than the computational

complexity $O(M^2)$ of constructing a full $M \times M$ k nearest neighbors (NN) similarity matrix, since $N \ll M$. The creation of the matrix $\mathbf{S}$ according to (5) has complexity $O(M^2)$ due to multiplication of the full matrix ($\mathbf{W}$) with diagonal matrices ($\mathbf{D}^{-1/2}$).

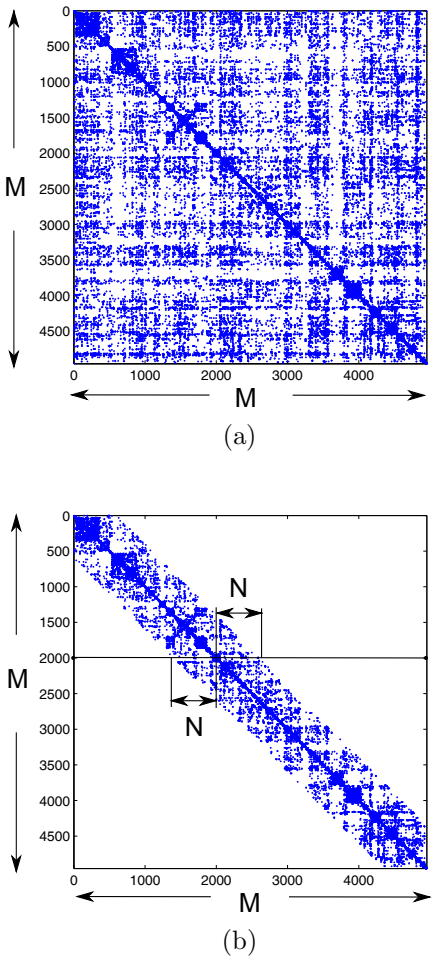

Figure 2: (a) Full (10-NN) Similarity Matrix, (b) Band (N=600) Similarity Matrix

Moreover, the label propagation solution (4) employing matrix ($\mathbf{I}-a\mathbf{S}$) inversion has complexity $O(M^3)$ [41, 42] and multiplication with the matrix $\mathbf{Y}$ has complexity $O(M^2Q)$. Solution (4) by inverting a band matrix using connectivity of Schur's complements [43], has complexity $O(M^2N) + O(M^2Q)$. The first complexity term refers to the inversion of the matrix ($\mathbf{I}-a\mathbf{S}$) whereas

the second one refers to the multiplication with matrix $\mathbf{Y}$.

Thus, we can conclude that the computational complexity of the proposed approach including the label propagation procedure is $O(M^2+M^2N+M^2Q+NM) \simeq O(M^2)$ which is much smaller than the full similarity matrix time complexity $O(M^2 + M^3 + M^2Q + M^2) \simeq O(M^3)$.

3.3. Constrained Label Propagation

Typically, label propagation techniques assume that facial images (or samples in general), which are similar to each other according to a similarity measure, should be assigned the same label.

In order to increase the facial image recognition accuracy using label propagation, information in the form of pairwise image similarity and dissimilarity constraints can be incorporated. In other words, data that satisfy similarity constraints should be assigned the same label and data that satisfy dissimilarity constraints should be assigned different labels [44]. Let S be the set of similar facial image pairs:

$$S = \{(i,j)|\mathbf{x}_i, \mathbf{x}_j \text{ must have the same label}\} \tag{11}$$

In our case, S contains the facial images that belong to the same facial image trajectory and most probably correspond to the same actor. Also let D be the set of dissimilar pairs:

$$D = \{(i,j)|\mathbf{x}_i, \mathbf{x}_j \text{ must have different labels}\} \tag{12}$$

In this paper, facial image pairs that appear on the same frame (and thus most probably belong to different actors) are included in D.

Two weight matrices $\mathbf{W}_s$, $\mathbf{W}_d$ can be constructed as follows:

$$W_{s,ij} = \begin{cases} 1, & \text{if } (i,j) \in S \\ 0, & \text{otherwise,} \end{cases} \tag{13}$$

$$W_{d,ij} = \begin{cases} 1, & \text{if } (i,j) \in D \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

The pairwise constraints can be diffused to neighboring nodes as follows. Let N_i be the neighborhood of the node i, based e.g., on thresholding the Euclidean distance between two nodes $\|\mathbf{x}_i - \mathbf{x}_j\| < e$ and $\mathbf{P} \in \Re^{M \times M}$ be the

sparse neighborhood probability matrix:

$$P_{ij} = \begin{cases} \frac{1}{|N_i|} & \text{if } j \in N_i \\ 0, & \text{otherwise,} \end{cases} \tag{15}$$

where $|N_i|$ is the cardinality of the set N_i. Pairwise similarity and dissimilarity information is propagated to neighboring nodes according to the iterative procedure:

$$\mathbf{F}_s^{(t)} = \alpha \mathbf{P} \mathbf{F}_S^{(t-1)} + (1-\alpha)\mathbf{W}_s, \tag{16}$$

$$\mathbf{F}_d^{(t)} = \alpha \mathbf{P} \mathbf{F}_D^{(t-1)} + (1-\alpha)\mathbf{W}_d, \tag{17}$$

where the parameter α, $0 \leq \alpha \leq 1$, controls the percentage of information which the node will receive from its neighbors and from the initial state. Iterative equations (16), (17) converge to the steady state solution [45]:

$$\mathbf{F}_s = (1-\alpha)(\mathbf{I} - \alpha \mathbf{P})^{-1} \mathbf{W}_s \tag{18}$$

$$\mathbf{F}_d = (1-\alpha)(\mathbf{I} - \alpha \mathbf{P})^{-1} \mathbf{W}_d. \tag{19}$$

As already stated, the incorporation of the pairwise similarity and dissimilarity constraints into label propagation aims at enhancing face recognition accuracy. This is obtained by the following procedure. First, the pairwise dissimilarity constraints between facial images that appear in the same video frame are taken into account by constructing the new weight matrix $\mathbf{W}_d$ as in (14). Respectively, the pairwise similarity constraints between the facial images that appear in the same facial image trajectory are taken into account by constructing the new weight matrix $\mathbf{W}_s$ as in (13). Then, the MLPP method, briefly described in Section 2, is carried out, in order to perform data dimensionality reduction by preserving the pairwise similarity and dissimilarity information. Then, label propagation is performed on the data projections, by incorporating the pairwise similarity and dissimilarity constraints to the objective function of label propagation, as follows:

$$\begin{aligned} Q(\mathbf{F}) = & \frac{1}{2} tr\left(\mathbf{F}^T \left(\sum_{k=1}^{K} \tau_k \mathbf{L_k} + \beta \mathbf{L_s} - \gamma \mathbf{L_d}\right) \mathbf{F}\right) + \\ & \mu tr((\mathbf{F} - \mathbf{Y})^T (\mathbf{F} - \mathbf{Y})) \end{aligned} \tag{20}$$

where $\mathbf{L}_k = \mathbf{D}_k - \mathbf{W}_k$ is the graph Laplacian for the k-th data represen-

tation 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 constrains, respectively. $\mathbf{L}_k$ varies according to the data representation, while $\mathbf{L}_s$, $\mathbf{L}_d$ are constant for all representations. Finally, $\mathbf{D}_s, \mathbf{D}_d$ and $\mathbf{D}_k$ are the diagonal degree matrices with entries $D_{s,ii} = \sum_{j=1}^{M} F_{s,ij}$, $D_{d,ii} = \sum_{j=1}^{M} F_{d,ij}$ and $D_{k,ii} = \sum_{j=1}^{M} W_{k,ij}$. The parameters β, γ are chosen in such a way that the matrix $\sum_{k=1}^{K} \tau_k \mathbf{L_k} + \beta \mathbf{L_s} - \gamma \mathbf{L_d}$ is positive definite. Minimization of equation (20) leads to the following label propagation solution:

$$\mathbf{F} = \mu \left(\alpha \mathbf{I} + \sum_{k=1}^{K} \tau_k \mathbf{L}_k + \beta \mathbf{L}_s - \gamma \mathbf{L}_d \right)^{-1} \mathbf{Y}. \tag{21}$$

This is the so-called CMLPP-CLP label propagation method. The merging of the constrained label propagation method with the pruning on the similarity matrix, i.e. the pruned label propagation (PB-MLPP-CLP), results to the proposed Pruned Constrained Band MLPP-CLP method using a band similarity matrix which will be called PCB-MLPP-CLP.

4. Experiments

4.1. Dataset and Method Application

Experimental evaluation of the proposed technique was performed on facial image label propagation in three stereoscopic Hollywood movies having a total duration of more than 6 hours and 528,348 frames in total. Person identity (label) propagation was performed on the facial images that appear in the two (left, right) video channels of these movies. Similar to [11] the input consisted of the pixel values of the facial image regions, after scaling them to the same dimensions of 41×31 pixels and converting them to a 1-D vector of 1271 dimensions. Dimensionality reduction according to MLPP was applied to these images in each channel separately. Thus, data dimensionality is reduced from 1271 to 75 dimensions. This is the dimension of $\mathbf{x}_i$ in (10). Then, label propagation is performed. For label propagation initialization, K-means clustering was used and only 5% of the facial images were manually labeled. As we have two ($K = 2$) different data representations on stereo video, namely the left and right stereo channels, late fusion [11] of the two data representations was performed. Both similarity and dissimilarity constraints were incorporated in the objective function of label propagation.

The band similarity matrix was computed according to (10). Since the total number of the extracted facial images using the face detector [46] and the single channel face tracker [47] is very large, experiments have been conducted with a small dataset of facial images from every movie. In total, 13850 images were selected from the three movies, which represent 5.85% of the detected/tracked facial images.

More precisely, if a facial image trajectory that results from face tracking contains less than 20 facial images then only the first facial image of the trajectory was selected. If the facial image trajectory contains more than 20 facial images, then every 10 facial images of the trajectory one facial image was selected for annotation (i.e., the 1st, 10th, 20th, etc.). The case where the different facial images that are contained in the same facial image trajectory are assigned different labels, practically does not exist due to the fact that these images will be placed in the similarity matrix in temporal order and thus will be neighboring ones. The propagation of similarity constraints to neighboring nodes (facial images) according to (18), will ensure that these images are assigned the same label.

Details about the dataset such as the number of actors classes, the number of facial images and the number of images which were initially labeled are given in Table 1.

Table 1: Dataset description

	Actors (classes)	Dataset size	Number of initially labeled images
Movie 1	27	5398	270
Movie 2	45	3498	175
Movie 3	59	4954	248
total	131	13850	693

4.2. Pruned Label Propagation Performance (without constraints)

In this section, we examine the effect of similarity matrix pruning on label propagation (PB-MLPP-CLP), measured by the obtained face recognition accuracy. It should be stressed that the constraints described in Section 3.3 are not included in the version of the proposed algorithm tested here (see Section 4.5 for the experimental evaluation of the pruned label propagation including constraints).

Figure 3 displays face recognition accuracy versus the percentage of the retained entries of the full (k-NN) similarity matrix $\alpha_p = \frac{2NM}{M^2} = \frac{2N}{M}\%$, where $2N$ denotes the width of the band of similarity matrix and M is the number of the facial images involved to the experiment ($M \times M$ being the dimensions of the similarity matrix). The horizontal lines show the classification accuracy of the full similarity matrix MLPP-CLP [11] method, which obviously does not depend on a_p. It should be noted here that the performance figures presented for MLPP-CLP in this figure as well as in the experiments presented in Sections 4.3 and 4.5 (e.g. in Table 2, Figure 8) are different from those presented in [11] because a different (better) initialization has been used. Figure 3 shows that the classification accuracy of the PB-MLPP-CLP pruned label propagation (without the constraints in Section 3.3) in one of the three movies (Movie 3) outperforms the classical MLPP-CLP method for most values of a_p most probably due to the fact that the similarity matrix pruning removes noise (semantically-unrelated facial images) from the graph which represent the similarity matrix. For the other two movies the proposed approach has equal (Movie 2) or very similar but inferior performance (Movie 1) to the MLPP-CLP method, for certain values of a_p. Moreover, we can notice that the classification accuracy for one movie (Movie 1), is increased as the percentage a_p increases. However, in Movies 2, 3, the classification accuracy decreases after a value of the a_p (namely $a_p = 0.15$ in Movie 2 and $a_p = 0.2$ in Movie 3). This can be attributed to the similarity matrix entries introduction which offers additional useful information until a point (value of a_p). After this point, the additional entries correspond to noise and as a result, the classification accuracy is decreased.

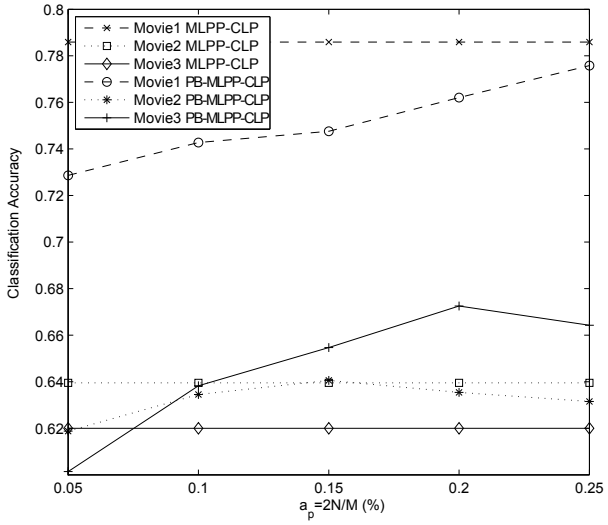

Figure 3: Face classification accuracy vs Approximation Percentage of retained similarity matrix entries.

Regarding computational complexity, let T_f, T_p be the execution time for the calculation of the full (k-NN) and the band similarity matrix, respectively. Figure 4 shows plots of the ratio $r_1 = \frac{T_f}{T_p}$ versus the percentage of the retained entries around the main diagonal (a_p) for the three movies. As the ratio r_1 is always bigger than one, pruning accelerates similarity matrix construction. Moreover, r_1 decreases towards 1 as a_p increases which is expected since as a_p tends to 1, the pruned similarity matrix tends to the full matrix. As can be observed in this figure the computational savings for a_p values that provide the best classification accuracy results, e.g. $a_p = 0.15, 0.2$ are significant (almost 6.94 and 5.36 times faster respectively). The plots follow very well the theoretical relation between r_1 and a_p, which is $r_1 = \frac{T_f}{T_p} = \frac{M^2}{2MN} = \frac{M}{2N} = \frac{1}{a_p}$.

Moreover, let T_{LP_f}, T_{LP_p} be the execution time of the label propagation procedure involving the full and the pruned similarity matrix as described in (4). Figure 5 shows the ratio $r_2 = \frac{T_{LP_f}+T_f}{T_{LP_p}+T_p}$, of the total execution times $T_{LP_f} + T_f$ and $T_{LP_p} + T_p$ versus the percentage of retained entries a_p. One can notice that the matrix construction and label propagation execution time is considerably smaller for the proposed pruning method. Indeed, in terms of computation time required for the entire label propagation procedure, including the construction time of the similarity matrix (5), the proposed

method is for example almost 4.07 and 3.84 times faster in Movie 3 than the full (k-NN) similarity matrix approach (MLPP-CLP) for values of $a_p = 0.15, 0.2$ in which the best recognition accuracy is presented. Moreover, it can be noticed that r_2 decreases as the percentage a_p is increased and tends to the value one. The ratio r_2 is slightly different between the movies due to the different number of manually assigned facial image labels in the algorithm initialization, which comprises the 5% of the detected/tracked facial images.

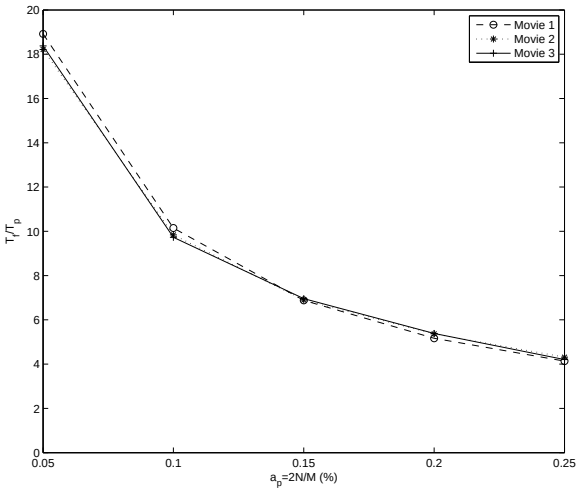

Figure 4: Ratio of similarity matrix construction time between method [11] and proposed pruning method versus the percentage a_p of the retained image similarity entries

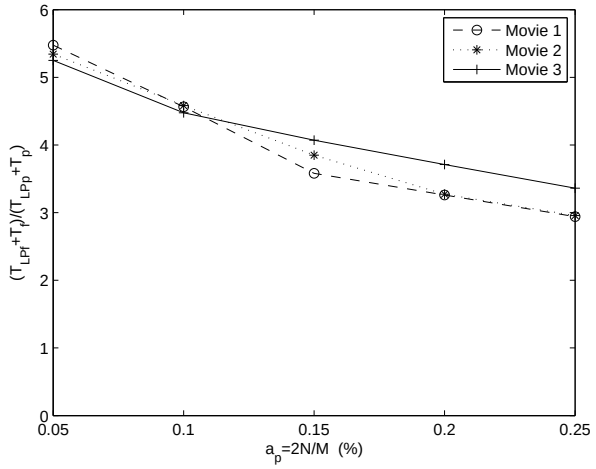

Figure 5: Ratio of similarity matrix construction time and label propagation execution time between method [11] and proposed pruning method versus the percentage a_p of the retained image similarity entries

The above experimental results show that face classification accuracy is not always increasing when one increases a_p but rather indicate that the behavior of the algorithm varies between videos. Thus, an easy and convenient conclusion of the form "the more entries you retain the larger the complexity and the face classification accuracy of the algorithm" cannot be reached. As a result, the choice of a_p shall be performed experimentally on each video, based on the label propagation computational complexity one can afford, in conjunction with the face classification errors he/she can tolerate. However, as a rule of thumb, a_p values in the range of $0.1 - 0.2$ work well in terms of both aspects (good classification, significant decrease of complexity).

4.3. Constrained Label Propagation Performance (without pruning)

In this section, we examine the effect of incorporating similarity and dissimilarity constraints in the objective function of label propagation procedure. More specifically, experimental results of the Constrained MLPP-CLP (CMLPP-CLP) without pruning, on the three stereo movies, when 5% of the facial images are manually labeled for initialization are shown in Table 2 in comparison to those of MLPP-CLP. Similarity constraints involve facial images in the same facial trajectory. All pairs of such images are included in set S in (11) since we assume that they depict the same person. Dissimilarity constraints involve facial images in the same frame. All pairs of such images

are included in set D in (12) since they obviously depict different person. We notice that incorporation of the pairwise constraints into the objective function of label propagation increases the classification accuracy on average by approximately 4.605%.

In addition, the performance of the proposed CMLPP-CLP method has been compared with that of other recent and older label propagation techniques on the three stereoscopic movies. The results are also presented in Table 2. Performance evaluation of these methods had been conducted by providing them with the same initial labelled images used by our algorithm. In order to be able to operate on the stereoscopic videos, the methods involved in the comparison were fed (for each movie) with a facial images similarity matrix that was calculated by fusing the similarity matrices from the two video channels (left and right) using the formula $W = \sum_k \tau_k W_k$ where τ_k are the weights evaluated by MLPP (see Section 2). For the performance evaluation of the MLAN method described in [28], we fed to the algorithm four sets of visual features from each facial image: colour moment (CM, 420 dimensions), GIST features (512 dimensions), HOG 3x3 features (420 dimensions), and local binary patterns (LBP, 1239 dimensions). As can be seen in this Table, the proposed CMLPP-CLP approach outperforms all other label propagation techniques.

Label Propagation Methods	*Movie1*	*Movie2*	*Movie3*
[9]	0.7541	0.5668	0.6364
LGC [10]	0.7310	0.5488	0.6356
LPP [38]	0.7426	0.5605	0.6150
NPE [48]	0.7564	0.5746	0.6314
OLPP [49]	0.6840	0.4868	0.6306
PCLPP [50]	0.7571	0.5977	0.6254
MAD [24]	0.7462	0.5631	0.6467
GoLPP [51]	0.6926	0.5952	0.5891
OMNI-Prop [26]	0.7502	0.5520	0.6501
CAMLP [27]	0.7308	0.5483	0.6322
MLAN [28]	0.5911	0.4380	0.4750
MLPP-CLP [11]	0.7859	0.6395	0.6200
CMLPP-CLP [proposed]	**0.8012**	**0.6722**	**0.7101**

Table 2: Comparison of CMLPP-CLP with other label propagation methods.

4.4. Choice of "constraints vs labels" Strategy

In an incremental semi-automatic label propagation task, the classification accuracy obtained, when a certain percentage of facial images is labeled, may be unsatisfactory. Then, a human annotator can perform two different actions towards reaching a desired classification accuracy: a) he can manually label additional unlabeled images or b) he can place additional pairwise facial image similarity or dissimilarity constraints. Thus the following questions naturally arise with respect to a "constraints vs labels" strategy: which of the two actions is more beneficial? What is the effect, in a certain label propagation problem of a) inserting one more constraint or b) labeling one more unlabeled image? Investigations towards answering these questions were conducted. Specifically, given a desired classification accuracy, the ratio of additional images that have to be manually labeled in order to achieve this accuracy, versus the required number of pairwise constraints needed to obtain the same accuracy was calculated. More specifically, let N_{RL} be the current number of manually labeled images, N_{TL} be the number of manually labeled images required in order to reach the desired classification accuracy P (without the use of any pairwise constraints) and N_c be the number of pairwise image similarity constraints needed (in addition to the N_{RL} labeled images) in order to reach P. For a given value of N_{TL} (and thus for a certain desired classification accuracy) and for a certain number of N_{RL}, the ratio r of the additional labeled images needed to reach P over the required pairwise constraints N_c needed to achieve the same target accuracy is given by: $r = (N_{TL} - N_{RL})/N_c$. Large values of r , e.g. values close to 1, mean that adding a labeling constraint has almost the same effect on the classification accuracy as labeling one more unlabeled image. On the contrary, small values of r denote that much more additional facial image constraints than labels are needed in order to reach the desired accuracy P.

For example, $r = 0.5$ means that, in order to reach P, one needs to place twice as many labeling constraints than additional labels on images in order to reach P. We have conducted such an experiment in one video, having 116 facial images. Plots of r versus N_{RL} for various values of N_{TL} are shown in Figure 6. Six curves are depicted in the Figure 6, each corresponds to a fixed N_{TL} value, which in turn corresponds to a certain classification accuracy value P. This figure suggests that r increases in general, as N_{RL} increases and never goes above 1. This means that, as the number of labeled images increases, the effect of the constraints increases. However, since r is in more cases significantly below 1, always less additional labeled images are needed

than additional labeling constraints to reach the desired accuracy.

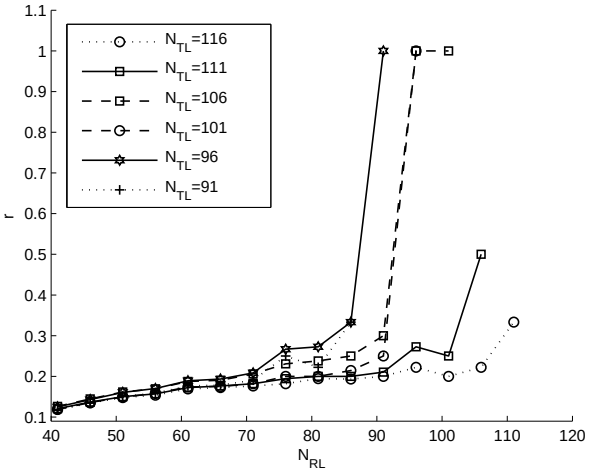

Figure 6: Ratio of the number additional labeled images over the number of additional labeling constraints for achieving desired facial image classification accuracy.

The above investigation naturally involves not only the effect of a "labels vs constraints" strategy towards reaching a certain (target) classification accuracy but also the cost/effort associated with each of the two actions (adding a label or a constraint). Usually, the effort of labeling an image is larger than that of assigning a pairwise labeling constraint. Indeed, when an annotator wants to annotate a facial image with one of the Q available labels (Q-class problem), in the worst case, he will have to check at least Q representative image samples (one per label) and (mentally) judge their similarity to the image to be labeled. In the best case, the annotator may remember all representative Q facial images and may need to make no image comparison at all for labeling an unlabeled image. Thus, on average, the annotator can take a labeling decision by examining/comparing $Q/2$ representative facial images with the image to be labeled. On the other hand, when the same annotator wants to place a pairwise image similarity constraint of the form "two images are similar" or "two images are dissimilar", then the decision is taken by just comparing the two images. Based on the above, the cost/effort for the annotator of assigning a label, denoted as C_l can be roughly set to $Q/2$ times the cost C_c of assigning a constraint: $C_l = (Q/2)C_c$. Having defined the cost for placing a new label or labeling constraint and assuming that the annotator

has already labeled N_{RL} images, the cost for reaching a desired accuracy by using only image labeling is $(N_{TL} - N_{RL})\frac{Q}{2}C_c$ whereas the cost of reaching the same accuracy through labeling constraints is $N_c C_c$. The ratio C of these costs is then:

$$C = \frac{(N_{TL} - N_{RL})\frac{Q}{2}C_c}{N_c C_c} = r\frac{Q}{2}. \quad (22)$$

Obviously, C values larger than one indicate that labeling is more laborious than inserting labeling constraints. Figure 7 shows plots of C versus N_{RL} for $Q = 46$ (the dataset contains 46 different classes/persons) for the previous experiment. As can be observed, C is larger than one for all values of N_{RL} and thus, labeling is more laborious than inserting constraints.

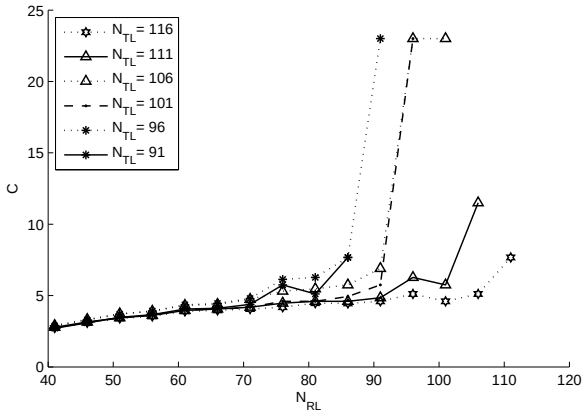

Figure 7: Ratio of the cost for reaching a desired accuracy by using only image labeling over the cost of reaching the same accuracy through labeling constraints.

4.5. Pruned Constrained Label Propagation Performance

In this section, we analyze the performance of the pruned constrained label propagation approach that includes the constraints in Section 3.3. In total 9,003 pairwise similarity and 2,034 pairwise dissimilarity constraints were automatically incorporated in the three movies. Figure 8 shows the obtained recognition accuracies of the proposed Pruned Constrained Band MLPP-CLP (PCB-MLPP-CLP) method and a) the MLPP-CLP method described in [11] and b) the proposed constrained MLPP-CLP (CMLPP-CLP) method i.e. the method that includes the constraints in Section III.C but utilizes the full (k-NN) matrix $\mathbf{W}$. Results are presented versus the percentage of the retained (band) entries of the similarity matrix $a_p = \frac{2NM}{M^2} = \frac{2N}{M}(\%)$.

The MLPP-CLP recognition accuracies form horizontal lines due to the fact that they are not depended by the percentage of the retained entries around the main diagonal (a_p). With respect to MLPP-CLP, we can notice that in two out of three movies (namely Movie 2,3), the PCB-MLPP-CLP method outperforms the classical MLPP-CLP method for most values of a_p (namely $a_p \geq 10\%$). This performance advantage is probably justified by the fact that pruning of the similarity matrix results in removing similarity entries that act as noise (semantically-unrelated facial images) in the label propagation procedure and also by the fact that the PCB-MLPP-CLP algorithm utilizes the additional information of the incorporated similarity and dissimilarity constraints. For Movie 1, one can notice that the MLPP-CLP recognition accuracy is significantly high, much higher than in the other two movies. This fact probably explains why in this movie the proposed PCB-MLPP-CLP performs worse than MLPP-CLP: data (facial images) quality is better in this movie (bigger facial images, more frontal faces), thus pruning entries of the similarity matrix removes useful information rather than denoising the data. Moreover, the removal of useful information can be noticed in Movie 2 as PCB-MLPP-CLP classification accuracy decreases above $a_p = 0.15$.

With respect to CMLPP-CLP (the non-pruned version of the constraints including proposed algorithm), one can observe in Figure 8 that PCB-MLPP-CLP has inferior but comparable performance. For large values of a_p (e.g. above 0.15) the performance of the two methods is very close and thus pruning does not significantly affect the performance of the algorithm.

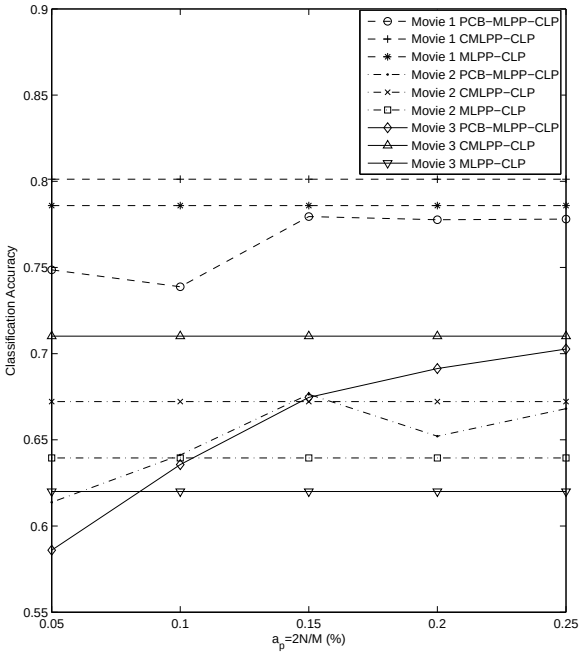

Figure 8: Face recognition accuracy vs the percentage of retained entries of the similarity matrix in PCB-MLPP-CLP. The MLPP-CLP [11] and CMLPP-CLP accuracies are constant (horizontal lines) since there is no pruning involved.

Figure 9 shows plots of the $r_3 = \frac{T_f}{T_p}$ where T_f is the time for the computation of the full (k-NN) similarity matrix of the MLPP-CLP method, T_p is the time for the computation of the pruned similarity matrix in the PCB-MLPP-CLP method. As in Section 4.2, when the ratio is bigger than one, we can observe a speedup. The r_3 decreases as the a_p is increased, which is expected because then the band similarity matrix tends to the full matrix. According to this plot, the computational savings for a_p values that provide best classification accuracy results, e.g. $a_p = 0.15, 0.25$ are significant (almost 6.8 and 4.14 times faster respectively).

Finally, let T_{LP_p} be the execution time of the pruned constrained label propagation involving the band similarity matrix and T_{LP_f} let be the execution time of the MLPP-CLP method [11] involving the full similarity matrix. The speedup ratio $r_4 = \frac{T_f + T_{LP_f}}{T_p + T_{LP_p}}$ in the execution time (matrix construction + label propagation) of the PCB-MLPP-CLP technique with respect to the

MLPP-CLP execution time (matrix construction + label propagation), is shown in Figure 10 for three movies.

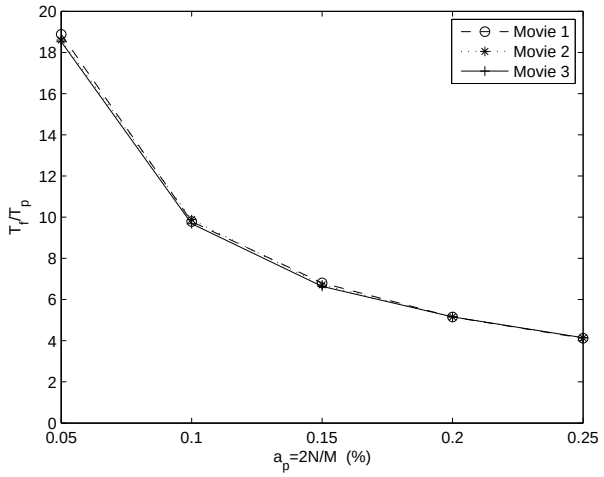

Figure 9: Speedup in the computation of the similarity matrix between MLPP-CLP and Pruned Constrained Band MLPP-CLP

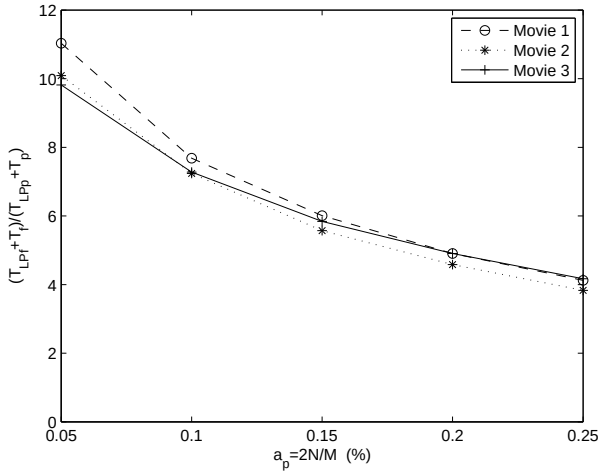

Figure 10: Speedup of label propagation and similarity matrix computation between MLPP-CLP and Pruned Constrained Band MLPP-CLP

Figure 10 shows that important speedup (up to 6 times faster in Movies 1,2 and up to 4.1 times faster for Movie 3) is obtained by the proposed

approach at the points $a_p = 0.15, 0.25$ in which the classification accuracy presents best values. Moreover, as expected, the speedup decreases as the percentage of the retained similarity matrix entries a_p increases and tends to one. Similar speedups were obtained by the PCB-MLPP-CLP pruned constrained approach over the full similarity matrix constrained CMLPP-CLP approach.

4.6. Performance in Monocular Videos

Tests with monocular sequences videos were also performed. The results for MLPP-CLP [11], PB-MLPP-CLP, CMLPP-CLP and PCB-MLPP-CLP on the right channel of Movie 2 are presented in Figure 11. A close look in this Figure reveals that the performance and behavior of the three proposed variants are similar to those exhibited in stereoscopic videos, as presented in the previous Sections. More specifically, the inclusion of constraints in the objective function (CMLPP-CLP) enhances the performance of the baseline algorithm (MLPP-CLP). Pruning of the similarity matrix (PB-MLPP-CLP) leads to a small decrease in the performance of MLPP-CLP, with a significant decrease in the computational complexity, as shown in the previous Sections. Finally, pruning combined with constraints inclusion (PCB-MLPP-CLP) leads to a performance better than that of MLPP-CLP but inferior to that of CMLPP-CLP, again with a reduced complexity compared to the latter approaches. By comparing Figure 11 with Figures 3, 8, that refer to stereoscopic videos, one can note that, as expected, the performance of both the baseline method and the proposed variants decreases when they are applied on monocular videos.

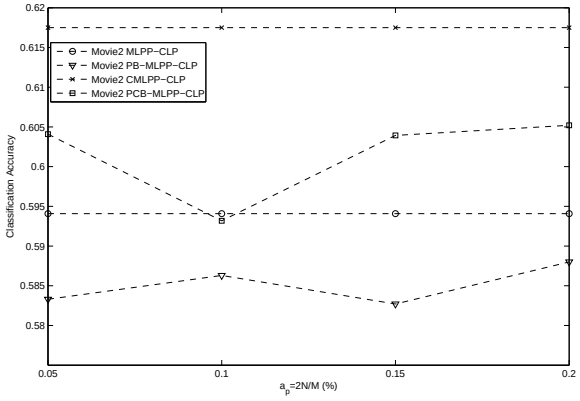

Figure 11: Face recognition accuracy vs the percentage of retained entries of the similarity matrix in a monocular video (right channel of Movie 2) for MLPP-CLP [11], PB-MLPP-CLP, CMLPP-CLP, PCB-MLPP-CLP. The MLPP-CLP and CMLPP-CLP accuracies are constant (horizontal lines) since there is no pruning involved.

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 which operates on multimedia data with multiple representations (but can easily be adopted to work on single representation data and data different than facial images), is called pruned label propagation and aims at accelerating the state of the art MLPP-CLP label propagation method [11]. Experiments on a large data set consisting of facial images extracted from three stereo movies show that a significant speedup is obtained by creating a band similarity matrix, which contains fewer pairwise facial image similarities. Such a speedup is also accompanied in many cases by an increase in the recognition accuracy as the similarity matrix pruning acts as denoising filter upon this matrix. Moreover, the paper proposes the use of similarity and dissimilarity labeling constraints in the objective function of label propagation which was shown to increase the classification accuracy of MLPP-CLP and outperform a number of recent and older label propagation approaches. The two approaches (pruning and inclusion of constraints) can be combined to obtain a faster and, in most cases, better performing (in terms of classification accuracy) version of MLPP-CLP. However the fact that the band similarity matrix creation requires the images to be temporally ordered limits

the applicability of the method on video data processed, for example, by a tracking algorithm. Nevertheless, since such data are generated in numerous cases, the proposed method has a quite broad range of applications.

Acknowledgment

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement number 316564 (IMPART). This publication reflects only the authors views. The European Union is not liable for any use that may be made of the information contained therein.

6. References

[1] O. M. Parkhi, A. Vedaldi, A. Zisserman, et al., Deep face recognition, Proceedings of the British Machine Vision Conference (BMVC) 1 (3) (2015) 6–17.

[2] I. Goodfellow, Y. Bengio, A. Courville, Y. Bengio, Deep learning, MIT press Cambridge, 2016.

[3] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, Advances in Neural Information Processing Systems (2012) 1097–1105.

[4] F. Schroff, D. Kalenichenko, J. Philbin, Facenet: A unified embedding for face recognition and clustering, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015) 815–823.

[5] X. Zhu, Semi-supervised learning literature survey, Tech. Rep. 1530, Computer Sciences, University of Wisconsin-Madison, http://www.cs.wisc.edu/∼jerryzhu/pub/ssl_survey.pdf (2005).

[6] O. Zoidi, A. Tefas, N. Nikolaidis, I. Pitas, Iterative label propagation on facial images, Proceedings of the 22nd European Signal Processing Conference (EUSIPCO) (2014) 1222–1226.

[7] O. Zoidi, E. Fotiadou, N. Nikolaidis, I. Pitas, Graph-based label propagation in digital media: A review, ACM Computing Surveys (CSUR) 47 (3) (2015) 48–83.

[8] F. Wang, C. Zhang, Label propagation through linear neighborhoods, IEEE Transactions on Knowledge and Data Engineering 20 (1) (2008) 55–67.

[9] X. Zhu, Z. Ghahramani, J. D. Lafferty, Semi-supervised learning using gaussian fields and harmonic functions, Proceedings of the 20th International conference on Machine learning (ICML-03) (2003) 912–919.

[10] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, B. Schölkopf, Learning with local and global consistency, Advances in Neural Information Processing Systems (2004) 321–328.

[11] O. Zoidi, A. Tefas, N. Nikolaidis, I. Pitas, Person identity label propagation in stereo videos, IEEE Transactions on Multimedia 16 (issue 5) (2014) 1358–1368.

[12] Z. Zhang, Y. Zhang, F. Li, M. Zhao, L. Zhang, S. Yan, Discriminative sparse flexible manifold embedding with novel graph for robust visual representation and label propagation, Pattern Recognition 61 (2017) 492–510.

[13] Z. Zhang, L. Jia, M. Zhang, B. Li, L. Zhang, F. Li, Discriminative clustering on manifold for adaptive transductive classification, Neural Networks 94 (2017) 260–273.

[14] L. Jia, Z. Zhang, Y. Zhang, Semi-supervised classification by nuclear-norm based transductive label propagation, International Conference on Neural Information Processing (2016) 375–384.

[15] L. Jia, Z. Zhang, W. Jiang, Transductive classification by robust linear neighborhood propagation, Pacific Rim Conference on Multimedia (2016) 296–305.

[16] L. Jia, Z. Zhang, L. Wang, W. Jiang, M. Zhao, Adaptive neighborhood propagation by joint L2, 1-norm regularized sparse coding for representation and classification, 2016 IEEE 16th International Conference on Data Mining (ICDM) (2016) 201–210.

[17] S. Yan, Y. Fu, T. Huang, B. Cheng, J. Yang, Learning with l1-graph analysis, IEEE Transactions on Image Processing 19 (4) (2010) 858–866.

[18] F. Dornaika, A. Bosaghzadeh, Adaptive graph construction using data self-representativeness for pattern classification, Information Sciences 325 (2015) 118–139.

[19] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, Y. Gong, Locality-constrained linear coding for image classification, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010 (2010) 3360–3367.

[20] F. Dornaika, M. T. Kejani, A. Bosaghzadeh, Graph construction using adaptive local hybrid coding scheme, Neural Networks 95 (2017) 91–101.

[21] W. Xiang, J. Wang, M. Long, Local hybrid coding for image classification, 2014 22nd International Conference on Pattern Recognition (ICPR) (2014) 3744–3749.

[22] F. Dornaika, R. Dahbi, A. Bosaghzadeh, Y. Ruichek, Efficient dynamic graph construction for inductive semi-supervised learning, Neural Networks 94 (2017) 192–203.

[23] F. Dornaika, A. Bosaghzadeh, H. Salmane, Y. Ruichek, Object categorization using adaptive graph-based semi-supervised learning, Handbook of Neural Computation (2017) 167–179.

[24] P. P. Talukdar, K. Crammer, New regularized algorithms for transductive learning, Joint European Conference on Machine Learning and Knowledge Discovery in Databases (2009) 442–457.

[25] S. Baluja, R. Seth, D. Sivakumar, Y. Jing, J. Yagnik, S. Kumar, D. Ravichandran, M. Aly, Video suggestion and discovery for youtube: taking random walks through the view graph, Proceedings of the 17th international conference on World Wide Web (2008) 895–904.

[26] Y. Yamaguchi, C. Faloutsos, H. Kitagawa, Omni-prop: Seamless node classification on arbitrary label correlation., Proceedings of the AAAI Conference on Artificial Intelligence (2015) 3122–3128.

[27] Y. Yamaguchi, C. Faloutsos, H. Kitagawa, Camlp: Confidence-aware modulated label propagation, Proceedings of the 2016 SIAM International Conference on Data Mining (2016) 513–521.

[28] F. Nie, G. Cai, X. Li, Multi-view clustering and semi-supervised classification with adaptive neighbours., Proceedings of the AAAI Conference on Artificial Intelligence (2017) 2408–2414.

[29] S. Kumar, M. Mohri, A. Talwalkar, Sampling techniques for the nystrom method, International Conference on Artificial Intelligence and Statistics (2009) 304–311.

[30] A. Talwalkar, S. Kumar, H. Rowley, Large-scale manifold learning, IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008 (2008) 1–8.

[31] K. Yu, S. Yu, V. Tresp, Blockwise supervised inference on large graphs, Proceedings of the 22nd ICML Workshop on Learning with Partially Classified Training Data.

[32] J. Tang, R. Hong, S. Yan, T.-S. Chua, G.-J. Qi, R. Jain, Image annotation by k nn-sparse graph-based label propagation over noisily tagged web images, ACM Transactions on Intelligent Systems and Technology (TIST) 2 (2) (2011) 14.

[33] C. Lehel, Z. B, Decomposition methods for label propagation, Proceedings of the International Conference on Knowledge Engineering, Principles and Techniques (2009) 1114.

[34] Y. Fujiwara, G. Irie, Efficient label propagation, Proceedings of the 31st International Conference on Machine Learning (ICML-14) (2014) 784–792.

[35] Z. Lu, L. Wang, Pairwise constraint propagation on multi-view data, arXiv preprint arXiv:1501.04284.

[36] X. Wang, B. Qian, I. Davidson, Labels vs. pairwise constraints: A unified view of label propagation and constrained spectral clustering, 2012 IEEE 12th International Conference on Data Mining (ICDM) (2012) 1146–1151.

[37] Z. Zhang, M. Zhao, T. W. Chow, Graph based constrained semi-supervised learning framework via label propagation over adaptive neighborhood, IEEE Transactions on Knowledge and Data Engineering 27 (9) (2015) 2362–2376.

[38] X. He, P. Niyogi, Locality preserving projections, Advances in neural information processing systems 16 (2004) 153–160.

[39] L. Van Der Maaten, E. Postma, J. Van den Herik, Dimensionality reduction: A Comparative Review, Journal of Machine Learning Research 10 (2009) 66–71.

[40] G. Yu, H. Peng, J. Wei, Q. Ma, Robust locality preserving projections with pairwise constraints, Journal of Computational Information Systems 6 (5) (2010) 1631–1636.

[41] J. B. Fraleigh, R. A. Beauregard, Linear algebra, Addison-Wesley Publishing Company, 1987.

[42] M. Belkin, P. Niyogi, V. Sindhwani, Manifold regularization: A geometric framework for learning from labeled and unlabeled examples, Journal of Machine Learning Research 7 (2006) 2399–2434.

[43] A. Mahmood, D. Lynch, L. Philipp, A fast banded matrix inversion using connectivity of schur's complements, IEEE International Conference on Systems Engineering, (1991) 303–306.

[44] O. Zoidi, N. Nikolaidis, I. Pitas, Semi-supervised dimensionality reduction on data with multiple representations for label propagation on facial images, Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2014) 6019–6023.

[45] G. H. Golub, C. F. Van Loan, Matrix computations 3, (2012), JHU Press.

[46] G. Stamou, M. Krinidis, N. Nikolaidis, I. Pitas, A monocular system for automatic face detection and tracking, Visual Communications and Image Processing 2005 5960 (2005) 794–802.

[47] O. Zoidi, A. Tefas, I. Pitas, Visual object tracking based on local steering kernels and color histograms, IEEE Transactions on Circuits and Systems for Video Technology 23 (5) (2013) 870–882.

[48] X. He, D. Cai, S. Yan, H.-J. Zhang, Neighborhood preserving embedding, 10th IEEE International Conference on Computer Vision, 2005. ICCV 2005 2 (2005) 1208–1213.

[49] L. Zhu, S. Zhu, Face recognition based on orthogonal discriminant locality preserving projections, Neurocomputing 70 (7-9) (2007) 1543–1546.

[50] H. Cevikalp, J. Verbeek, F. Jurie, A. Klaser, Semi-supervised dimensionality reduction using pairwise equivalence constraints, VISAPP'08 - 3rd International Conference on Computer Vision Theory and Applications 1 (2008) 489–496.

[51] L. Zhang, L. Qiao, S. Chen, Graph-optimized locality preserving projections, Pattern Recognition 43 (6) (2010) 1993–2002.