INCREMENTAL LABEL PROPAGATION ON FACIAL IMAGES

Efstratios Kakaletsis, Anastasios Tefas, Nikos Nikolaidis, Ioannis Pitas

Department of Informatics , AIIA Lab
Aristotle University of Thessaloniki
GR-54124 Thessaloniki, GREECE
Emails:{ekakalets, tefas, nnik, pitas}@csd.auth.gr

ABSTRACT

The increasing computational complexity of label propagation-based facial image annotation when applied on multimedia data whose cardinality increases over the time (e.g., when analyzing video or movie content on-line), can be reduced by using an incremental approach. In this paper, a method for incremental label propagation on facial images is described. The similarity matrix is incrementally constructed by employing the kd-tree nearest neighbor algorithm. Furthermore, the matrix inversion, which is included in the label propagation solution, is calculated with a block-wise inversion formula involving the Woodbury matrix identity. Experiments show significant computational savings when the incremental approach is applied on a dataset of three full length movies. Moreover, the classification accuracy was improved in most cases.

Index Terms— Incremental label propagation, facial image annotation, incremental similarity matrix construction, blockwise inversion formulae, Woodbury matrix identity.

1. INTRODUCTION

Annotation of facial images can be done using label propagation techniques with the prior knowledge of some annotated facial images and by spreading the name labels from the set of labeled images to the unlabeled ones, employing a diffusion process while maintaining local and global labeling consistency [1], [2]. In the field of label propagation, or, more generally speaking, in graph-based semi-supervised learning [3], [4], many tasks employ datasets that evolve over time. Such a task is the annotation (tagging) of facial images [5] (derived from a face detector and a face tracker that produce "trajectories" of facial images, consisting of regions of interest (ROIs)) with names in a video stream that becomes available on-line, i.e. it is not known in its entirety. In this case, classical facial image annotation via label propagation till time instance $t+1$ must be done from scratch rather than using the results of the same operation already performed at time instance t. However, in such cases, incremental label propagation techniques can be used in order to avoid the increasing computational complexity of classical label propagation [1].

Various methods implementing incremental label propagation exist in the literature [6–9]. In [6], the proposed dynamic label propagation method (DLP) aims at solving semi-supervised multiclass and multi-label label propagation by combining the graph similarities and the label correlations in a dynamic way, while preserving the intrinsic structure of input data. Moreover, propagation of semantic labels in indoor sequences based on feature matching and incremental model update, is described in [7]. Furthermore, *LabelRankT* [8] is an online distributed algorithm for detecting, incrementally updating and monitoring, through label propagation, communities [10] in large-scale dynamic networks whose evolution generates huge realtime data streams. Finally, label propagation for detecting community structures in complex networks in realtime while dealing with the network changes incrementally is described in [9].

In this paper, a novel incremental label propagation approach is presented, which is based on Multiple-graph Locality Preserving Projections - Cluster-based Label Propagation (MLPP-CLP) algorithm [1]. The main aim is the speedup of MLPP-CLP method using the concept of incremental label propagation. To this end, we split the video time domain in intervals $n_tT, n_t = 1, \ldots, N_t$, the total duration being N_tT seconds. We start with the first video interval by performing face detection/tracking and we manually label an initial number of facial images (namely 5% of the facial images contained in the time interval). Label propagation is conducted by spreading the labels from the manually labeled facial images to the unlabeled ones in the same time interval. Subsequently, this process is repeated for consecutive time intervals and the label inference is conducted in an incremental way as the additional facial images are being detected/tracked on line, by: a) updating in every step the respective facial image similarity matrix $\mathbf{W}$ with the additional pairwise facial im-

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) and the European Union's Horizon 2020 research and innovation programme under grant agreement number 731667 (MULTIDRONE).

978-1-7281-0824-7/19/$31.00 ©2019 IEEE

age similarities and b) calculating the label propagation solution with the inversion of the matrix $\mathbf{I} - a\mathbf{S}$ in an incremental block-wise manner based on the Woodbury matrix identity [11]. This approach reduces computational complexity as these two operations are the most time-consuming ones in the label propagation process. Incremental facial label propagation continues till the video stream (or movie) comes to its end.

The rest of this paper is organized as follows. Since the proposed method is an extension of the MLLP-CLP method [1], it is briefly described in Section 2. Section 3 describes the details of the proposed incremental label propagation method. In Section 4, we present the datasets and the experiments which have been conducted to measure the facial label propagation accuracy and the reduction in computational complexity. Finally, conclusions are presented in Section 5.

2. MLPP-CLP FACIAL IMAGE LABEL PROPAGATION

Assume a set of labeled facial images $X_L = \{\mathbf{x}_i, i = 1, \ldots, m_l\}$ which have been assigned labels (actor/person names) from the label set $L = \{l_j, j = 1, \ldots, Q\}$ and a set of unlabeled facial images $X_U = \{\mathbf{x}_i, i = 1, \ldots, m_u\}$. Their union is given by $X = \{\mathbf{x}_1, ..., \mathbf{x}_{m_l}, \mathbf{x}_{m_l+1}, ..., \mathbf{x}_M\}, M = m_l + m_u$ [2]. The objective of label propagation is to spread the facial image labels from the set of the labeled images X_L to the set of the unlabeled images X_U, while maintaining local and global facial image labeling consistency [2]. The initial information about the labeled data is described by the $M \times Q$ matrix $\mathbf{Y}$, defined as:

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

The algorithm begins with the construction of a symmetric facial image similarity matrix $\mathbf{W}$ which represents the facial image similarity graph, as label propagation should be performed across similar facial images. The rows/columns of the matrix correspond to the temporally ordered facial images. 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 the graph nodes. This similarity is computed according to the heat kernel equation:

$$W_{ij} = e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma}} \quad (2)$$

where σ is the mean edge length distance among neighbors. The construction of such a matrix has computational complexity and memory requirements of the order $O(M^2)$, even if a k nearest neighbor (NN) matrix [1] is constructed.

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

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

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 degree matrix having entries $D_{ii} = \sum_j W_{ij}$ and μ is a regularization parameter. This minimization problem leads to the following solution [1]:

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

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

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

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

$$y_i = \underset{j \in 1, \ldots, Q}{\arg\max} [f_j^1, \ldots, f_j^M]. \quad (6)$$

The regularization framework (3) can be easily extended to the case of label propagation on multiview (e.g. stereo) 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). In this case, the regularization framework (3) takes the form:

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

subject to the constraint:

$$\sum_{k=1}^{K} \tau_k = 1, \quad (8)$$

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

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

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 Cluster-based Label Propagation (MLPP-CLP) was introduced in [1], being a variant of the Locality Preserving Projections (LPP) method [12]. It performs dimensionality reduction [13] 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 [1]. The weight τ_k of each data representation in the construction of the projection matrix $\mathbf{A}$ is optimal for the label propagation cost function (7), given that the data feature extraction was performed according to MLPP. More details about the method can be found in [1].

3. INCREMENTAL FACIAL IMAGES LABEL PROPAGATION

The proposed Incremental Facial Images Label Propagation (ILP) is a variant of the MLPP-CLP method [1] described in Section 2, with incremental calculation of similarity matrix $\mathbf{W}^{(n_t)}$ and incremental matrix $\mathbf{I}-a\mathbf{S}^{(n_t)}$ inversion by employing a block-wise matrix inversion formula.

Assume that we start with M initial facial images in the set $X_M = \{\mathbf{x}_1,\dots,\mathbf{x}_M\}$ corresponding to time interval $[0, n_t T]$. We obtain m new labeled and unlabeled facial images samples $X_{in} = \{\mathbf{x}_{M+1},\dots,\mathbf{x}_{M+m}\}$ that are derived by on-line face detection and tracking (and manual annotation of a small number of images) over the period $[n_t T, (n_t+1)T]$, resulting in a new image data set $X_{M+m} = \{\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_{m_l},\dots,\mathbf{x}_M,\mathbf{x}_{M+1},\dots,\mathbf{x}_{M+m}\}$. For every facial image $\mathbf{x}_i \in X_{in}$, we then search the k-nearest samples $\mathbf{x}_p \in N(\mathbf{x}_i)$ using the kd-tree algorithm [14], calculate the incremental facial image similarity matrix $\mathbf{W}^{(n_t+1)}$, update the matrix $\mathbf{S}^{(n_t+1)}$ and invert matrix $\mathbf{I}-a\mathbf{S}^{(n_t+1)}$ in order to conduct label inference. The objective is to spread the facial labels in the unlabeled facial images of the new facial image set X_{in} using the the manually labeled images of X_{in} as well as those in X_M. Note however that labels assigned to images in X_M can also change, i.e., the approach is equivalent to performing label propagation in X_{M+m} in a non-incremental way. Including a small set of labelled images in each new time interval ensures that appearances of new persons, not present so far in the video, will be correctly handled. If all persons to be labelled appear in the first interval, manual labeling in each new interval can be omitted.

3.1. Incremental Similarity Matrix Calculation

Assume that $\mathbf{W}^{(n_t)} \in \Re^{M\times M}$ is the similarity matrix for the time interval $[0, n_t T]$ that represents the similarities among M facial images. Then, the facial image similarity matrix $\mathbf{W}^{(n_t+1)} \in \Re^{(M+m)\times(M+m)}$ for the time interval $[0,(n_t+1)T]$ is given by:

$$\mathbf{W}^{(n_t+1)} = \begin{bmatrix} \mathbf{W}^{(n_t)} & \mathbf{W}' \\ \mathbf{W}'^T & \mathbf{W}_m \end{bmatrix}, \tag{10}$$

where $\mathbf{W}_m \in \Re^{m\times m}$ is the matrix with the pairwise facial image similarities between the new facial image entries in X_{in} (detected/tracked in the time interval $[n_t T, (n_t+1)T]$). $\mathbf{W}' \in \Re^{M\times m}$ is the matrix with the pairwise facial image similarities between the new/additional m facial image entries and (previous) M facial image entries, which had already been used in the similarity matrix $\mathbf{W}^{(n_t)}$. It should be noted that $\mathbf{W}'$, is found using the nearest image neighbors according to a kd-tree algorithm [14].

3.2. Incremental Inversion of the Matrix I-aS

Assume that $\mathbf{D}^{'}_{(n_t)}$, $\mathbf{D}^{'}_m$ are the diagonal degree matrices having entries $D'^{ii}_{(n_t)} = \sum_j W^{(n_t)}_{ij} + \sum_j W'_{ij} = D^{ii}_{(n_t)} + \sum_j W'_{ij}$ and $D'^{ii}_{m} = \sum_j W_{m_{ij}} + \sum_i W'_{ij}$, respectively. The diagonal degree matrix $\mathbf{D}_{(n_t+1)}$ for the time interval $[0,(n_t+1)T]$ is given by:

$$\mathbf{D}_{(n_t+1)} = \begin{bmatrix} \mathbf{D}^{'}_{(n_t)} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{'}_m \end{bmatrix} \tag{11}$$

Thus, the matrix $\mathbf{S}^{(n_t+1)}$ is given by the following equation:

$$\mathbf{S}^{(n_t+1)} = \mathbf{D}^{-1/2}_{(n_t+1)} \mathbf{W}^{(n_t+1)} \mathbf{D}^{-1/2}_{(n_t+1)} = = \begin{bmatrix} \mathbf{D}^{'-1/2}_{(n_t)} \mathbf{W}^{(n_t)} \mathbf{D}^{'-1/2}_{(n_t)} & \mathbf{D}^{'-1/2}_{(n_t)} \mathbf{W}' \mathbf{D}^{'-1/2}_m \\ \mathbf{D}^{'-1/2}_m \mathbf{W}'^T \mathbf{D}^{'-1/2}_{(n_t)} & \mathbf{D}^{'-1/2}_m \mathbf{W}_m \mathbf{D}^{'-1/2}_m \end{bmatrix} = \begin{bmatrix} \mathbf{S}^{'(n_t)} & \mathbf{S}^{'} \\ \mathbf{S}^{'T} & \mathbf{S}^{'}_m \end{bmatrix}, \tag{12}$$

The matrix $\mathbf{I}-a\mathbf{S}^{(n_t+1)}$ which appears in the solution (4) or (9) is computed as:

$$\mathbf{P}^{(n_t+1)} = \mathbf{I}-a\mathbf{S}^{(n_t+1)} = \begin{bmatrix} \mathbf{V} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}, \tag{13}$$

where $\mathbf{V} = \mathbf{I} - a\mathbf{S}^{'(n_t)}, \mathbf{B} = -a\mathbf{S}', \mathbf{C} = \mathbf{B}^T = -a\mathbf{S}'^T, \mathbf{D} = \mathbf{I} - a\mathbf{S}^{'}_m$ are matrix sub-blocks of size $M\times M$, $M\times m$, $m\times M$ and $m\times m$, respectively and $\mathbf{I}$ the identity matrix. The inversion of the matrix $\mathbf{P}^{(n_t+1)}$ is computed according to a block-wise inversion formula using the Woodbury matrix identity [11]:

$$(\mathbf{P}^{(n_t+1)})^{-1} = (\mathbf{I}-a\mathbf{S}^{(n_t+1)})^{-1} = \begin{bmatrix} \mathbf{V} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = = \begin{bmatrix} \mathbf{V}^{-1} + \mathbf{V}^{-1}\mathbf{B}\mathbf{Z}^{-1}\mathbf{C}\mathbf{V}^{-1} & -\mathbf{V}^{-1}\mathbf{B}\mathbf{Z}^{-1} \\ -\mathbf{Z}^{-1}\mathbf{C}\mathbf{V}^{-1} & \mathbf{Z}^{-1} \end{bmatrix}, \tag{14}$$

where $\mathbf{Z} = (\mathbf{D} - \mathbf{C}\mathbf{V}^{-1}\mathbf{B})$. $\mathbf{V}$ is a square matrix, therefore it can be inverted. Furthermore, $\mathbf{V}$ and $\mathbf{Z}$ are nonsingular matrices [15].

3.3. Computational complexity study

In non-incremental label propagation using (4), the creation of the matrix $\mathbf{S}$ according to (5) has complexity $O((M+m)^2) \simeq O(M^2)$ for $m \ll M$, due to multiplication of the full matrix $\mathbf{W}$ with diagonal matrices $\mathbf{D}^{-1/2}$. Moreover, the non-incremental construction of the similarity matrix has computational complexity and memory requirements of the order $O((M+m)^2) \simeq O(M^2)$ even if an k nearest neighbor (NN) matrix is constructed. Finally, the label propagation solution (4) using MLPP-CLP and employing matrix $\mathbf{I}-a\mathbf{S}$ inversion has complexity $O((M+m)^3) \simeq O(M^3)$ [16, 17] and multiplication with the matrix $\mathbf{Y}$ has complexity $O((M+m)^2 Q) \simeq O(M^2 Q)$. As a result, the non-incremental approach has time complexity $O(2M^2 + M^3 + M^2 Q) \simeq O(M^3)$.

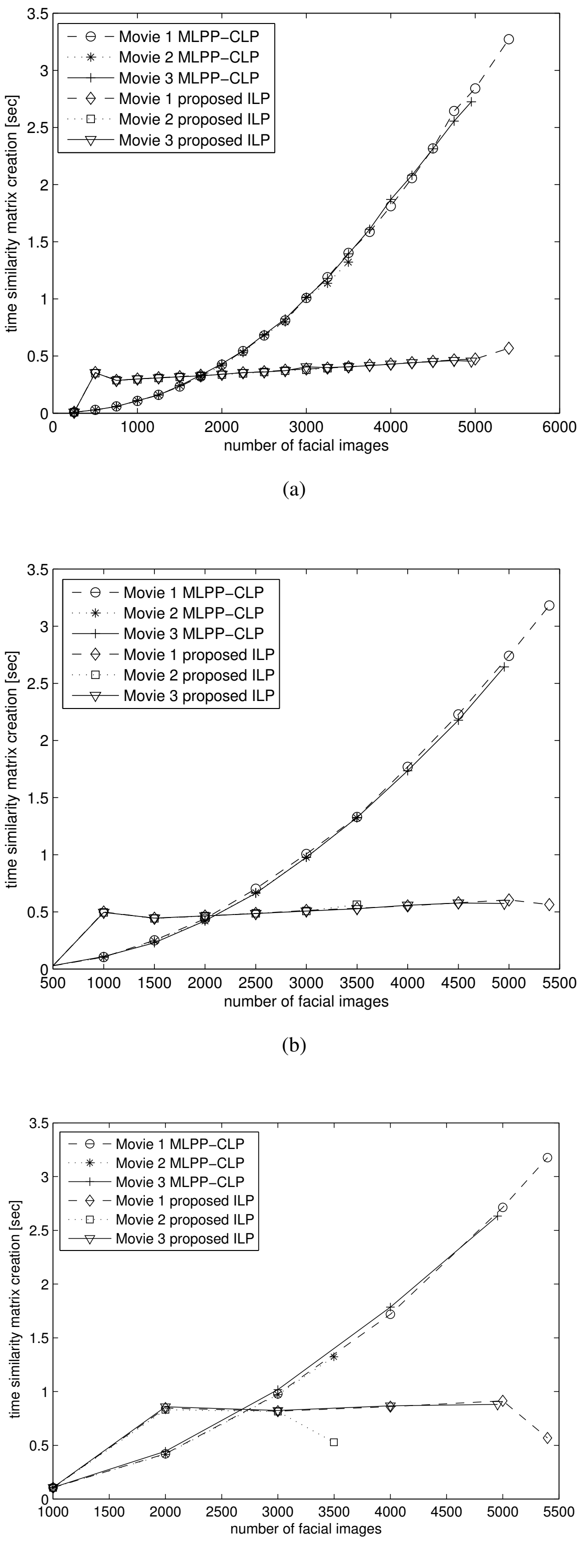

Fig. 1: Execution time for similarity matrix calculation of MLPP-CLP and proposed ILP method versus the number of processed facial images $k \cdot m$ for a) $m = 250$, b) $m = 500$, c) $m = 1000$, respectively and the three movies.

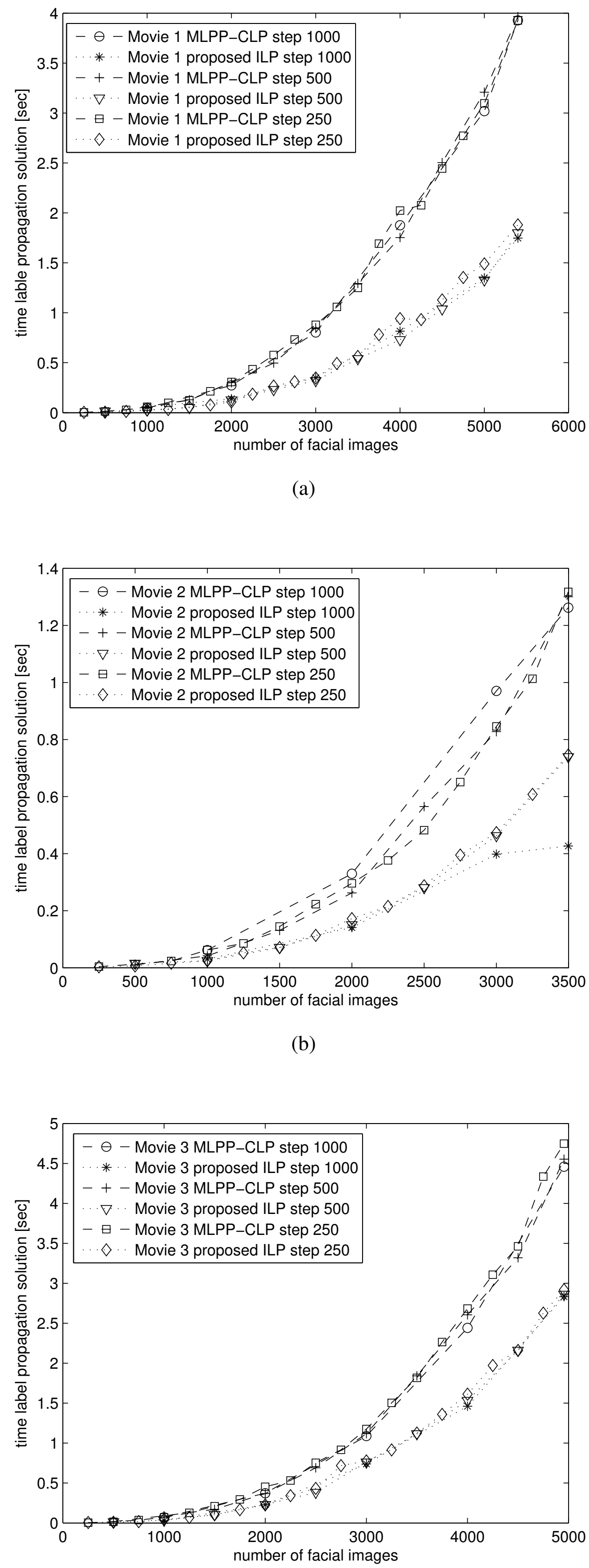

Fig. 2: Execution time for label propagation solution of the MLPP-CLP method and proposed ILP technique (excluding similarity matrix calculation) versus the number of processed facial images $k \cdot m$ for a) Movie 1, b) Movie 2, c) Movie 3, respectively and various values of m.

However, the construction of the incremental similarity matrix of the ILP method has time complexity $O(m^2) + O(2Mm) \simeq O(Mm)$ for $m \ll M$. Furthermore, the creation of the incremental matrix $\mathbf{S}^{(n_t+1)}$ according to (12) has complexity $O(m^2) + O(2Mm) \simeq O(Mm)$. Finally, it can be shown that the algorithm using blockwise inversion for inverting a matrix $(M+m) \times (M+m)$ runs with the same time complexity as the matrix multiplication that is used in (14) [18]. As there are matrix multiplication algorithms with a complexity of $O(M^{2.3727})$ this will also be the complexity of the block-wise matrix inversion (14). Thus, the proposed iterative label propagation method (4) has time complexity $O((M+m)^{2.3727}) \simeq O(M^{2.3727})$ due to incremental inversion of the matrix $\mathbf{I} - a\mathbf{S}^{(n_t+1)}$ plus $O((M+m)^2 Q) \simeq O(M^2 Q)$ due to multiplication with matrix $\mathbf{Y}$ in (4). As a result, the incremental label propagation has time complexity $O(2Mm + M^{2.3727} + M^2 Q) \simeq O(M^{2.3727} + M^2 Q)$ which is much less than that of the MLPP-CLP time complexity of $O(M^3)$.

4. EXPERIMENTAL EVALUATION

Experiments on incremental facial image label propagation (ILP) have been conducted on three full length stereoscopic movies having a total duration of more than 6 hours (528348 frames). The facial images contained in these movies were derived by applying face detection [19] and tracking [20]. Due to the fact that the number of detected/tracked frames is very large, only 5.85% of them have been used for annotation with incremental label propagation (namely 5398, 3498, 4954 in every movie respectively).

Every facial image ROI is of size 41×31 pixels. In an approach slightly different from what it has been described before, each movie is segmented into unequal time intervals, each containing the same number of facial images m, $m = 250, 500, 1000$. For every segment, manual labeling of 0.05m (i.e. 5%) of the facial images in the segment was performed.

The computational complexity experiments focused on: a) incremental similarity matrix $\mathbf{W}$ construction b) incremental label propagation (4), (9), which contains a block-wise matrix $\mathbf{I} - a\mathbf{S}$ inversion. These experiments measured the execution time of MATLAB code running on a PC equipped with an Intel Xeon CPU at 2.40 GHz with 72GB RAM and 64-bit Windows Operating System.

4.1. Classification Accuracy Performance

As facial label propagation essentially performs face classification, Table 1 illustrates the classification accuracy of the proposed ILP method and the respective MLPP-CLP [1] results. The classification accuracy of ILP is evaluated in the final time interval, thus it refers to the whole facial image dataset, i.e. when the label propagation is completed. ILP method outperforms MLPP-CLP in most cases. More precisely, the classification accuracy gains of 2.5% (on average) that were observed for the various m, could be attributed to the difference in the calculation of the similarity matrix in ILP and MLPP-CLP. In MLPP-CLP, the k nearest neighbors of the facial images are evaluated from within the entire set of facial images available at a certain time instance. In the case of ILP however, as can be seen in (10) $\mathbf{W}^{(n_t+1)}$ is essentially an approximation of the "full" similarity matrix involved in MLPP-CLP. For example, evaluation of sub-matrix $\mathbf{W}_m$ involves finding the nearest neighbors only between the new m facial images. This approximation seems to improve the classification accuracy.

Although the proposed method aimed at improving MLPP-CLP that operates on multi-view (stereoscopic in our case) data, it can be easily applied on monocular videos as well. Table 2 presents the recognition accuracy of ILP and MLPP-CLP when applied on monocular videos. Similar gains can be observed.

Table 1: ILP Recognition Accuracy Performance (stereoscopic movies)

	m	ILP	MLPP-CLP
Movie1	1000	**0.846**	0.8189
	500	0.8315	**0.8326**
	250	**0.8351**	0.8121
Movie2	1000	**0.7288**	0.7094
	500	**0.7365**	0.7090
	250	**0.7381**	0.6990
Movie3	1000	**0.7242**	0.6942
	500	**0.7156**	0.6797
	250	**0.7053**	0.6841

Table 2: ILP Recognition Accuracy Performance (monocular movies)

	m	ILP	MLPP-CLP
Movie1	1000	**0.8239**	0.8147
	500	0.8287	**0.8292**
	250	0.8200	**0.8248**
Movie2	1000	**0.6736**	0.6531
	500	**0.7088**	0.6527
	250	**0.7107**	0.6635
Movie3	1000	**0.7043**	0.6767
	500	**0.6951**	0.6863
	250	**0.6962**	0.6853

4.2. Execution Time

Figure 1 shows the similarity matrix construction time for both methods, namely MLPP-CLP [1] and the proposed ILP technique for each stereoscopic movie and for $m = 250, 500, 1000$. The horizontal axis shows the total number of facial images $k \cdot m, (k = 1, 2, ...)$ at a certain time instance. We observe a speedup by a factor of 2.5 to 5.58 for $m = 1000$, 2.35 to 5.7 for $m = 500$ and 3.2 to 5.98 for $m = 250$, respectively, when we performed facial label propagation for the entire movie (last point in each curve).

The differences in speedup between the 3 movies are due to the different number of faces in each movie. The bigger the number of images, the larger the speed up. Figure 2 shows label propagation solution execution time (excluding similarity matrix calculation) for the MLPP-CLP and the ILP methods for the three stereoscopic movies separately. We observe a speedup of 2.55 in Movie 1, 2.95 in Movie 2 and 1.66 in Movie 3 respectively for the $m = 250, 500, 1000$, when we perform label propagation for the entire movie.

5. CONCLUSION

In this paper, a novel incremental method for propagating person identity labels on facial images extracted from stereo or monocular videos was introduced. The main aim was the reduction of the computational complexity of the state of the art MLPP-CLP [1] method, by performing the similarity matrix construction $\mathbf{W}$ and matrix $\mathbf{I}$-a$\mathbf{S}$ inversion in an incremental way. Experiments on a data set consisting of facial images extracted from three monocular and stereoscopic movies show that a significant speedup is obtained by using incremental label propagation. Moreover, the classification accuracy was improved in most cases. It should be noted that the proposed approach can be also used for speeding up label propagation in other applications where data to be labelled are evolving over time.

6. REFERENCES

[1] O. Zoidi, A. Tefas, N. Nikolaidis, and I. Pitas, "Person identity label propagation in stereo videos," *IEEE Transactions on Multimedia*, vol. 16, no. issue 5, pp. 1358–1368, 2014.

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

[3] Z. Ghahramani, "Graph-based semi-supervised learning," *Department of Engineering University of Cambridge, UK*, 2012.

[4] X. Zhu and A. B. Goldberg, "Introduction to semi-supervised learning," *Synthesis lectures on artificial intelligence and machine learning*, vol. 3, no. 1, pp. 1–130, 2009.

[5] A.Nithya and K.Haridas, "An overview on facial image annotation," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 2, no. 7, pp. 5158–5163, 2014.

[6] B. Wang, Z. Tu, and J. K. Tsotsos, "Dynamic label propagation for semi-supervised multi-class multi-label classification," *IEEE International Conference on Computer Vision (ICCV)*, pp. 425–432, 2013.

[7] J. Rituerto, A. Murillo, and J. Košecka, "Label propagation in videos indoors with an incremental non-parametric model update," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2383–2389, 2011.

[8] J. Xie, M. Chen, and B. K. Szymanski, "Labelrankt: Incremental community detection in dynamic networks via label propagation," *Workshop on Dynamic Networks Management and Mining*, pp. 25–32, 2013.

[9] S. Pang, C. Chen, and T. Wei, "A realtime community detection algorithm: incremental label propagation," *First International Conference on Future Information Networks*, pp. 313–317, 2009.

[10] M. E. Newman, "Detecting community structure in networks," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 38, no. 2, pp. 321–330, 2004.

[11] M. A. Woodbury, "Inverting modified matrices," *Memorandum Rept. 42, Statistical Research Group*, p. 4, 1950.

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

[13] L. J. van der Maaten, E. O. Postma, and H. J. van den Herik, "Dimensionality reduction: A comparative review," *Journal of Machine Learning Research*, vol. 10, no. 1-41, pp. 66–71, 2009.

[14] J. L. Bentley, "K-d trees for semidynamic point sets," *Sixth annual Symposium on Computational Geometry*, pp. 187–197, 1990.

[15] D. S. Bernstein, *Matrix mathematics: theory, facts, and formulas*. Princeton university press, 2009.

[16] J. B. Fraleigh and R. A. Beauregard, "Linear algebra. addison," *Wesley Publishing Co., Reading, MA*, p. 477, 1987.

[17] Belkin, Mikhail, Niyogi, Partha, and S. Vikas, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *Journal of Machine Learning Research*, pp. 7:2399–2434, 2006.

[18] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms 3rd ed*, 2009.

[19] G. Stamou, M. Krinidis, N. Nikolaidis, and I. Pitas, "A monocular system for automatic face detection and tracking," *Visual Communications and Image Processing 2005*, vol. 5960, pp. 794–802, 2005.

[20] 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, 2013.