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

Introduction

- Annotation (tagging) of facial images with names can be done through label propagation by using a small set of annotated facial images and spreading the name labels from this set to the unlabeled ones.
- Annotation of facial images in a video stream that becomes available on-line is a task where data evolve over time.
- 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 obtained at t.
- In such cases incremental label propagation can be used to reduce computational complexity.
- A novel incremental label propagation approach is presented, aiming to speed-up Multiple-graph Locality Preserving Projections - Cluster-based Label Propagation (MLPP-CLP) algorithm [1].

[1] O. Zoidi, A. Tefas, N.Nikolaidis, I.Pitas, *Person identity label propagation in stereo videos*, IEEE Transactions on Multimedia, vol 16(5), pp: 1358-1368, 2014.

Method Overview

- We split the video in intervals $n_tT, n_t = 1,\ldots,N_t$.
- We perform face detection/tracking in the first video interval and manually label a number of facial images (e.g. 5%) in it.
- Label propagation is conducted in this interval and the process is repeated for consecutive time intervals in an incremental way by:
  - updating in every step the respective facial image similarity matrix $\mathbf{W}$ with the additional pairwise similarities
  - calculating the propagation solution by inverting matrix $\mathbf{I}-a\mathbf{S}$ in an incremental block-wise manner based on the Woodbury matrix identity.

MLPP-CLP

Assume:

- a set of labeled facial images $X_L = \{\mathbf{x}_i, i = 1,\ldots,m_l\}$
- a label set $L = \{l_j, j = 1,\ldots,Q\}$
- a set of unlabeled facial images $X_U = \{\mathbf{x}_i, i = 1,\ldots,m_u\}$
- their union $X = \{\mathbf{x}_1,...,\mathbf{x}_{m_l},\mathbf{x}_{m_l+1},...,\mathbf{x}_M\}, M = m_l + m_u$

Information about labeled data is given by matrix $\mathbf{Y}$:

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

A symmetric facial image similarity matrix $\mathbf{W}$ 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 $\mathbf{F} = [\mathbf{f}_1^T,...,\mathbf{f}_M^T]^T \in R^{M\times Q}$.
$\mathbf{F}$ is calculated by solving a minimization problem leading to the following solution:

$$\mathbf{F} = (1-a)(\mathbf{I}-a\mathbf{S})^{-1}\mathbf{Y}, \quad (2)$$

where $\mathbf{S} = \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$, and $\mathbf{D}$ is the diagonal degree matrix, $D_{ii} = \sum_j W_{ij}$.
A facial image label is assigned to facial image i according to:

$$y_i = \arg\max_{j\in 1,\ldots,Q}[f_j^1,\ldots,f_j^M]. \quad (3)$$

MLPP-CLP (Cont'd)

The approach can be extended to multiview (e.g. stereo) facial images or data with K representations:

$$\mathbf{F} = (1-a)\left(\mathbf{I} - a\sum_k \tau_k\mathbf{S}_k\right)^{-1}\mathbf{Y}. \quad (4)$$

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 and providing dimensionality reduction (MLPP-CLP) was introduced in [1].

Incremental Label Propagation

Assume:

- M initial facial images in the set $X_M = \{\mathbf{x}_1,\ldots,\mathbf{x}_M\}$ in time interval $[0, n_tT]$
- m new labeled and unlabeled facial images in interval $[n_tT,(n_t+1)T]$ $X_{in} = \{\mathbf{x}_{M+1},\ldots,\mathbf{x}_{M+m}\}$

The result is a new image data set $X_{M+m} = \{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_{m_l},\ldots,\mathbf{x}_M,\mathbf{x}_{M+1},..,\mathbf{x}_{M+m}\}$.
The incremental similarity matrix is:

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

$\mathbf{W}_m$: matrix with pairwise similarities between the new facial images in X_{in}.
$\mathbf{W}'$: matrix with pairwise similarities between the new m facial image entries and (previous) M facial images, already used in $\mathbf{W}^{(n_t)}$.
Incremental matrix $\mathbf{S}^{(n_t+1)}$ appearing in (2), (4):

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

where $\mathbf{D}_{(n_t+1)} = \begin{bmatrix} \mathbf{D}_{(n_t)}' & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_m' \end{bmatrix}$ and $\mathbf{D}_{(n_t)}'\mathbf{D}_m'$ are appropriate diagonal matrices.
Incremental block-wise inversion of matrix $\mathbf{I} - a\mathbf{S}^{(n_t)}$ in (2), (4), based in the Woodbury matrix identity:

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

where $\mathbf{V} = \mathbf{I} - a\mathbf{S}'^{(n_t)}, \mathbf{B} = -a\mathbf{S}', \mathbf{C} = \mathbf{B}^T = -a\mathbf{S}'^T, \mathbf{D} = \mathbf{I} - a\mathbf{S}_m'$, $\mathbf{Z} = (\mathbf{D} - \mathbf{C}\mathbf{V}^{-1}\mathbf{B})$

Computational Complexity

Computational complexity per video interval:
Non-Incremental (MLPP-CLP):

- Similarity matrix construction: $O((M+m)^2) \simeq O(M^2)$ for $m \ll M$
- Propagation solution (2), (4): $O(2M^2+M^3+M^2Q) \simeq O(M^3)$

INCREMENTAL LP:

- Similarity matrix construction: $O(m^2) + O(2Mm) \simeq O(Mm)$
- Propagation solution (2), (4): $O(M^{2.3727} + M^2Q)$

Experimental Evaluation

- Experiments on three stereoscopic movies of total duration of more than 6 hours. Facial images were derived by face detection and tracking.
- A subset of the detected/tracked facial images (5398, 3498, 4954 for the 3 movies respectively) has been used in the experiments.
- Each movie is segmented into unequal intervals, each containing the same number of images m, $m = 250, 500, 1000$. 5% of the images in each interval are manually labeled.
- 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 $m = 250$ in similarity matrix construction was observed for the 3 movies, depending on the facial images in each movie.
- A speedup of 2.55 in Movie 1, 2.95 in Movie 2 and 1.66 in Movie 3 was observed in label propagation solution execution (for all m values).
- Classification accuracy gains of 2.5% (on average) were observed due to the difference in the calculation of the similarity matrix in the incremental approach.

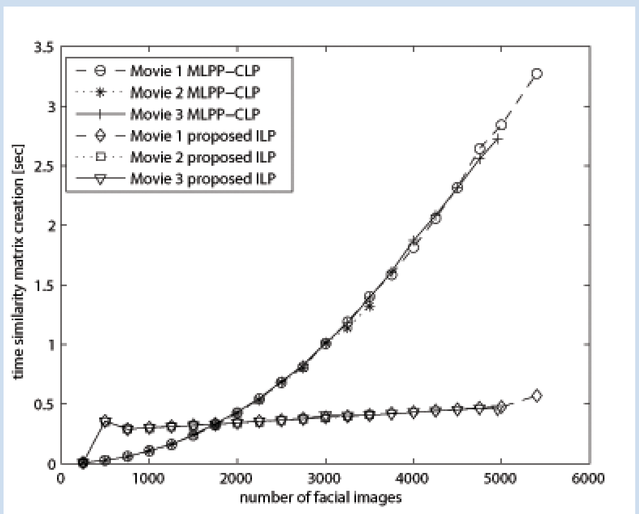

Fig.1 Similarity matrix calculation time for $m = 250$

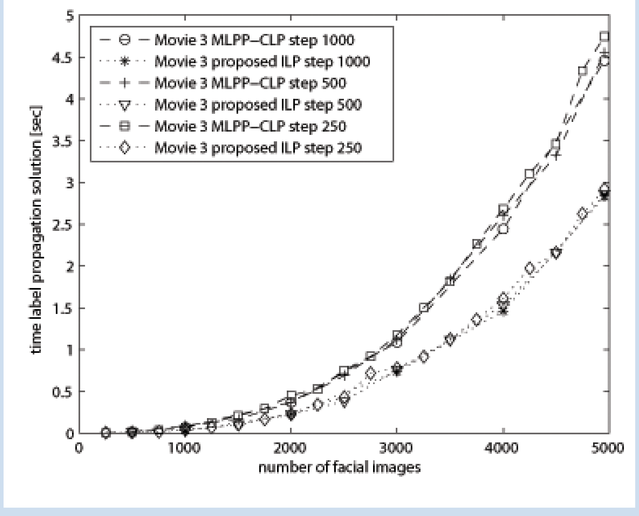

Fig.2 Label propagation solution time for movie 3.

TABLE I: ILP Recognition Accuracy Performance (stereoscopic movies)

	m	ILP	MLPP-CLP
Movie1	1000	**0.8209**	0.8189
	500	0.8002	**0.8326**
	250	0.8001	**0.8121**
Movie2	1000	0.6929	**0.7094**
	500	**0.7162**	0.7090
	250	**0.7035**	0.6990
Movie3	1000	0.6933	**0.6942**
	500	0.6791	**0.6797**
	250	0.6636	**0.6841**

Conclusions

- An incremental method for propagating person identity labels on facial images extracted from stereo, but also monocular, videos was introduced.
- A significant speedup is obtained. The classification accuracy was also improved in most cases.
- The proposed approach can be also used for speeding up label propagation in other applications where data are evolving over time.

Acknowledgements

The research leading to these results has received funding from the European Union Seventh Framework Programme under grant agreement no 316564 (IMPART) and the European Union's Horizon 2020 research and innovation programme under grant agreement no 731667 (MULTIDRONE).