Artificial Intelligence and Information Analysis (AIIA) Lab
Department of Informatics
Aristotle University of Thessaloniki, GREECE

ICIP 2018

Label Propagation on Facial Images Using Similarity and Dissimilarity Labelling Constraints

Efstratios Kakaletsis, Olga Zoidi, Ioannis Tsingalis,
Anastasios Tefas, **Nikos Nikolaidis**, Ioannis Pitas
{tefas, nikolaid, pitas}@aiia.csd.auth.gr

October 10, 2018

Introduction

- Tagging facial regions with person ID in images and videos is useful for archival/search but time consuming when done manually.
- One approach that can be used is label propagation:
  - Manual labeling of faces with person ID in specific video frames or images
  - Spreading the labels from the (small) labeled facial image dataset to the unlabeled images.
  - Semi-supervised classification approach

Label Propagation Example

Figure: Propagate the labels from the manually labelled images (in rectangles) to the remaining ones.

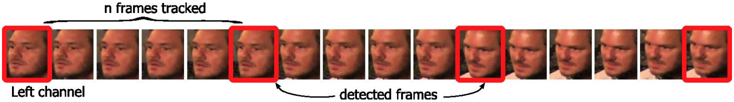
Overview

- **Goal:**
  Enhance classification performance (labeling accuracy) of the Multiple-graph Locality Preserving Projections – Cluster based Label Propagation (MLPP-CLP) technique (Zoidi et al [1]), when applied on facial images derived from stereo videos,
- **How:**
  Incorporate pairwise facial image similarity and dissimilarity constraints into the objective function of MLPP-CLP.

[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.

MLPP-CLP : Short Description

- Facial images are extracted by applying face detection and tracking in the left/ right view of a stereo video.
- Facial image trajectories are derived: sequences of facial images representing a tracked face over time.



- Each such facial trajectory is represented by one image (short trajectories) or more images (longer trajectories).

MLPP-CLP : Short Description

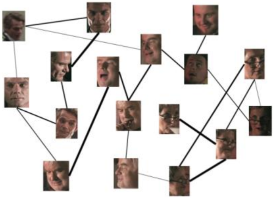
Inputs:

- Set of labeled facial images: $X_L = \{\mathbf{x}_i\}_{i=1}^{m_l}$
- Set of persons names (labels): $L = \{l_j\}_{j=1}^{Q}$
- Set of unlabeled data: $X_U = \{\mathbf{x}_i\}_{i=1}^{m_u}$
- Set of labeled and unlabeled facial images: $X = \{\mathbf{x}_1, ..., \mathbf{x}_{m_l}, \mathbf{x}_{m_l+1}, ..., \mathbf{x}_M\}, M = m_l + m_u$

Objective: spread the labels in L from the set of labeled data X_L to the set of unlabeled data X_U.

MLPP-CLP : Short Description

- MLPP-CLP: Extension of the Zhou et al[2] approach to data with multiple representations
- Facial images in stereoscopic video: K=2 data representations, left/right channel
- Build facial images similarity matrix **W** using heat kernel
- $W_{ij} = e^{-\frac{\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2}{\sigma}}, i \neq j, \mathbf{x}_i, \mathbf{x}_j \text{ are k-NN}$
- **W** represents the corresponding similarity graph (nodes=images)



[2] D. Zhou, O. Bousquet, T.N. Lal, J. Weston, and B. Scholkopf, "Learning with local and global consistency," NIPS 2004.

MLPP-CLP : Short Description

- Build matrix **Y** containing information regarding the labels in labeled data set X_L :

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

- Label Inference: Assign a score for every label to each facial image through matrix **F**:

$$\mathbf{F} = [\mathbf{f}_1^T, ..., \mathbf{f}_M^T]^T \in \mathbf{R}^{M \times Q}$$

  - F_{ij}: score for j-th label in i-th image
  - Q: number of labels, M: number of images

MLPP-CLP : Short Description

- $\mathbf{F}$ is found by solving a minimization problem which leads to the following solution:

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

- $\mathbf{S} = \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$
- $D_{ii} = \sum_j W_{ij}$ degree matrix
- Label y_i for i-th image: $y_i = \arg\max_{j \in 1,\ldots,Q} [f_{i1}, \ldots, f_{ij}, \ldots, f_{iQ}]$
  - Image is assigned the label with the highest score

MLPP-CLP : Short Description

- MLPP-CLP extends this approach to data with multiple representations,
- A separate graph is constructed for each of the K facial image representations
  - $K = 2$ for stereoscopic images: left / right view.
- Each graph is represented by the corresponding similarity matrix $\mathbf{W}_k, k = 1, ..., K$
- The regularization framework takes the form:

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

- $\mathbf{L}_k = \mathbf{D}_k - \mathbf{W}_k$: graph Laplacian for the k-th data representation.
- τ_k, $k = 1,...,K$: weight for the k-th data representation

MLPP-CLP : Short Description

- This leads to the following solution for **F**:
  - $\mathbf{F} = (1 - a)\left(\mathbf{I} - a \sum_k \tau_k \mathbf{S}_k\right)^{-1} \mathbf{Y}$
  - $\mathbf{S}_k = \mathbf{D}^{-1/2}\mathbf{W}_k\mathbf{D}^{-1/2}$
- MLPP also performs dimensionality reduction by extending Locality Preserving Projections (LPP) method[3] to a multiple-graph framework
- A single projection matrix **A** is constructed for all data representations, while preserving locality information and similarity/dissimilarity constraints.

[3] X. He, P. Niyogi, "Locality Preserving Projections", NIPS 2003

Proposed method: Constrained MLPP-CLP

- Proposed Constrained MLPP-CLP (CMLPP-CLP) approach incorporates pairwise image similarity and dissimilarity constraints in the MLPP-CLP objective function.
  - Similar images shall be assigned the same label:
    - S: set of similar facial image pairs:

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

    - S contains facial images belonging to the same facial image trajectory
  - They depict the same actor

Proposed method: Constrained MLPP-CLP

- Dissimilar images shall be assigned different labels:
  - D: set of dissimilar pairs:

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

  - D includes facial image pairs that appear on the same frame
    - They belong to different actors.

Proposed method: Constrained MLPP-CLP

- Two weight matrices $\mathbf{W}_s$, $\mathbf{W}_d$ are constructed:

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

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

Proposed method: Constrained MLPP-CLP

- Similarity and dissimilarity information is propagated to neighboring nodes according to an iterative procedure that converges to the steady state solution:

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

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

- $\mathbf{P} \in \Re^{M\times M}$: sparse neighborhood probability matrix:

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

- N_i: neighborhood of node i

Proposed method: Constrained MLPP-CLP

- Then dimensionality reduction is performed through MLPP.
- Label propagation is conducted on the data projections, by incorporating the pairwise similarity and dissimilarity constraints to the label propagation objective function:

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

- $\mathbf{L}_s = \mathbf{D}_s - \mathbf{F}_s$, $\mathbf{L}_d = \mathbf{D}_d - \mathbf{F}_d$: graph Laplacians of the similarity and dissimilarity constrains.

Proposed method: Constrained MLPP-CLP

- Minimization of $Q(\mathbf{F})$ leads to the following solution for $\mathbf{F}$:

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

Experimental evaluation

- Dataset:
  - 3 stereo movies
  - Duration: 2 hours each
  - Facial images per movie: 5300, 3500, 5000 (after retaining one/more image(s) per facial trajectory)
  - Actors (classes) per movie: 26, 44, 58.
- 5% of the facial images were manually labeled.
- Dimensionality reduction down to 75 dimensions

Experimental evaluation

	MLPP-CLP	CMLPP-CLP
Movie 1	0.7859	**0.801223**
Movie 2	0.6395	**0.672213**
Movie 3	0.62	**0.710133**

Classification accuracy obtained using MLPP-CLP and Constrained MLPP-CLP (CMLPP-CLP).

- Incorporation of pairwise constraints into the objective function of label propagation increases the classification accuracy by **4.6%** on average.
- Recent experiments showed that CMLPP-CLP outperforms both older and recent approaches
  - OMNI-Prop, Yamaguchi et al., AAAI 2015
  - CAMLP, Yamaguchi et al., SIAM Int. Conf. on Data Mining, 2016
  - MLAN, Nie et al, AAAI 2017

Choice of ”constraints vs labels” strategy

- A human annotator can perform two different actions towards reaching a desired classification accuracy:
  - Manually label additional unlabeled images or
  - Place additional pairwise facial image similarity or dissimilarity constraints
- Experiments were conducted in order to answer the following questions:
  - What is the effect of inserting one or more constraints or labeling one or more images?
  - Which of the two actions is more beneficial?

Choice of ”constraints vs labels” strategy

- N_{RL}: current number of manually labeled images
- N_{TL}: number of manually labeled images required in order to reach the desired classification accuracy P (without using pairwise constraints)
- N_c: number of pairwise similarity constraints needed (in addition to the N_{RL} labeled images) in order to reach P
- Ratio r of additional labeled images over additional constraints for achieving desired classification accuracy:

$$r = (N_{TL} - N_{RL})/N_c$$

- Small r (below 1): more constraints than labels are needed in order to reach the desired accuracy P

Choice of ”constraints vs labels” strategy

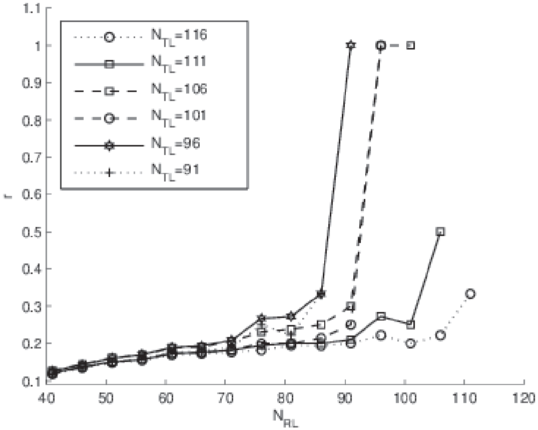


Figure: r versus N_{RL} for various values of N_{TL} (desired accuracy P).

Choice of ”constraints vs labels” strategy

- r is in most cases significantly below 1:
  - Less labeled images than labeling constraints are neededto reach the desired accuracy
  - Labeled images carry more information than constraints.
- However, the effort of labeling an image is larger than that of assigning a pairwise labeling constraint.

Conclusions

- A novel method (CMLPP-CLP) for propagating person identity labels on facial images extracted from stereo videos was introduced.
- It incorporates similarity and dissimilarity labelling constraints in order to increase the classification accuracy
- The proposed method outperforms current methods.
- It can be used to perform label propagation in other types of images or data in general.
- It can be easily adapted to work with images taken from monocular cameras.
- An investigation of labels vs constraints strategy that should be followed in order to reach a desired accuracy was also conducted.

Conclusions

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).

Thank you for your attention!