Identity inference: generalizing person re-identification scenarios

Svebor Karaman   Andrew D. Bagdanov

MICC - Media Integration and Communication Center
University of Florence
svebor.karaman@unifi.it, bagdanov@dsi.unifi.it

1st International Workshop on Re-Identification
12 October 2012

Outline of the talk

Introduction

Re-identification

Recognition of an individual at different times, over different camera views and/or locations, and considering a large number of candidate individuals

- Re-identification performance is usually evaluated as a **retrieval problem**.
  Good for descriptor performance analysis but limited wrt the final real world application where the problem should be considered as a labelling problem.
- **Re-identification protocols** may be **ambiguous**, and evolved protocols are not generalization of the simpler ones.
  We will formally define all the standard scenarios and their natural generalization.

Related work

- The majority of existing research on the person re-identification problem has concentrated on the **development of sophisticated features for describing the visual appearance** of targets [Schwartz and Davis, 2009, Gray and Tao, 2008, Bak et al., 2011, Farenzena et al., 2010, Cheng et al., 2011, Cai and Pietikäinen, 2011, Bazzani et al., 2012]
- Less research on classification or ranking technique. Ensemble RankSVM [Prosser et al., 2010] which learns a ranking SVM model to solve the single-shot re-identification problem.
- How to solve efficiently re-identification as a labelling problem? CRFs for multi-target tracking [Yang and Nevatia, 2012]

Identity inference as generalization of re-identification

Given a set of labels (individuals): $\mathcal{L} = \{1, \ldots N\}$, and images of individuals from $\mathcal{L}$ detected in a video collection:

$$\mathcal{I} = \{x_i \mid i = 1 \ldots D\}.$$

Each image x_i being represented by a feature vector $\mathbf{x}_i \equiv \mathbf{x}(x_i)$, the corresponding label is given by $y_i \equiv y(x_i)$.

Identity inference as generalization of re-identification

Given a set of labels (individuals): $\mathcal{L} = \{1, \ldots N\}$, and images of individuals from $\mathcal{L}$ detected in a video collection:

$$\mathcal{I} = \{x_i \mid i = 1 \ldots D\}.$$

Each image x_i being represented by a feature vector $\mathbf{x}_i \equiv \mathbf{x}(x_i)$, the corresponding label is given by $y_i \equiv y(x_i)$.

Re-identification problem definition

A re-identification problem is a tuple $\mathcal{R} = (\mathcal{X}, \mathcal{Z})$ completely characterized by its gallery and test image sets ($\mathcal{X}$ and $\mathcal{Z}$, respectively).
A solution to an instance of a re-identification problem is a mapping from the test images $\mathcal{Z}$ to the set of all permutations of $\mathcal{L}$.

Identity inference as generalization of re-identification

Gallery images

Gallery images $\mathcal{X}$, defined as:

$$\mathcal{X} = \{\mathcal{X}_j \mid j = 1 \dots N\}, \text{where } \mathcal{X}_j \subset \{x \mid y(x) = j\}$$

Identity inference as generalization of re-identification

Gallery images

Gallery images $\mathcal{X}$, defined as:

$$\mathcal{X} \quad = \quad \{\mathcal{X}_j \mid j = 1 \ldots N\}\,, \text{where } \mathcal{X}_j \subset \{x \mid y(x) = j\}$$

Test images

The set of test images $\mathcal{Z}$, defined as:

$$\mathcal{Z} = \{\mathcal{Z}_j \mid j = 1 \ldots M\} \subset \mathcal{P}(\mathcal{I})$$

$\mathcal{P}$ is the powerset operator (i.e. $\mathcal{P}(I)$ is the set of all subsets of $\mathcal{I}$). For all $\mathcal{Z}_j \in \mathcal{Z}$:

- $x, x' \in \mathcal{Z}_j \Rightarrow y(x) = y(x')$, sets in $\mathcal{Z}$ have homogeneous labels
- and $\mathcal{Z}_j \in \mathcal{Z} \Rightarrow \mathcal{Z}_j \cap \mathcal{X}_i = \emptyset, \forall i \in \{1 \ldots N\}$, disjoint test and gallery sets

Re-identification scenarios

- **Single-versus-all (SvsAll)**: single gallery image for each individual, *all remaining instances* of each individual as test. $\mathcal{R}_{\text{SvsAll}} = (\mathcal{X}, \mathcal{Z})$:

$$\begin{aligned} \mathcal{X}_j &= \{x\} \text{ for some } x \in \{x \mid y(x) = j\}, \text{ and} \\ \mathcal{Z}_j &= \{\{x\} \mid x \in \mathcal{I} \setminus \mathcal{X}_j \text{ and } y(x) = j\} \end{aligned}$$

Re-identification scenarios

- **Single-versus-all (SvsAll)**: single gallery image for each individual, *all remaining instances* of each individual as test. $\mathcal{R}_{\text{SvsAll}} = (\mathcal{X}, \mathcal{Z})$:

$$
\begin{aligned}
\mathcal{X}_j &= \{x\} \text{ for some } x \in \{x \mid y(x) = j\}, \text{ and} \\
\mathcal{Z}_j &= \{\{x\} \mid x \in \mathcal{I} \setminus \mathcal{X}_j \text{ and } y(x) = j\}
\end{aligned}
$$

- **Multi-versus-single shot (MvsS)**: G gallery images of each person, each of the test sets $\mathcal{Z}_j$ contains a single image. $\mathcal{R}_{\text{MvsS}} = (\mathcal{X}, \mathcal{Z})$:

$$
\begin{aligned}
\mathcal{X}_j &\subset \{x \mid y(x) = j\} \text{ and } |\mathcal{X}_j| = G \ \forall j \text{ and} \\
\mathcal{Z}_j &= \{x\} \text{ for some } x \notin \mathcal{X}_j \text{ s.t. } y(x) = j.
\end{aligned}
$$

Re-identification scenarios

- **Single-versus-all (SvsAll)**: single gallery image for each individual, *all remaining instances* of each individual as test. $\mathcal{R}_{\text{SvsAll}} = (\mathcal{X}, \mathcal{Z})$:

$$\begin{aligned} \mathcal{X}_j &= \{x\} \text{ for some } x \in \{x \mid y(x) = j\}, \text{ and} \\ \mathcal{Z}_j &= \{\{x\} \mid x \in \mathcal{I} \setminus \mathcal{X}_j \text{ and } y(x) = j\} \end{aligned}$$

- **Multi-versus-single shot (MvsS)**: G gallery images of each person, each of the test sets $\mathcal{Z}_j$ contains a single image. $\mathcal{R}_{\text{MvsS}} = (\mathcal{X}, \mathcal{Z})$:

$$\begin{aligned} \mathcal{X}_j &\subset \{x \mid y(x) = j\} \text{ and } |\mathcal{X}_j| = G \ \forall j \text{ and} \\ \mathcal{Z}_j &= \{x\} \text{ for some } x \notin \mathcal{X}_j \text{ s.t. } y(x) = j. \end{aligned}$$

- **Multi-versus-multi shot (MvsM)**: the gallery and test sets of each person both have G images. $\mathcal{R}_{\text{MvsM}} = (\mathcal{X}, \mathcal{Z})$:

$$\begin{aligned} \mathcal{X}_j &\subset \{x \mid y(x) = j\} \text{ and } |\mathcal{X}_j| = G \ \forall j \text{ and} \\ \mathcal{Z}_j &\subset \{x \mid y(x) = j \text{ and } x \notin \mathcal{X}_j\} \text{ and } |\mathcal{Z}_j| = G \ \forall j. \end{aligned}$$

Identity inference

Identity inference problem definition

Having *few* labeled images, label *many* unknown images without explicit knowledge that groups of images represent the same individual.

- The formulation of the *single-versus-all* re-identification falls within the scope of identity inference
- Neither the multi-versus-single nor the multi-versus-multi formulations are a generalization of this case to multiple gallery images

Identity inference

Identity inference problem definition

Having *few* labeled images, label *many* unknown images without explicit knowledge that groups of images represent the same individual.

- The formulation of the *single-versus-all* re-identification falls within the scope of identity inference
- Neither the multi-versus-single nor the multi-versus-multi formulations are a generalization of this case to multiple gallery images

Identity inference formulation

A *multi-versus-all* configuration, $\mathcal{R}_{\text{MvsAll}} = (\mathcal{X}, \mathcal{Z})$:

$$\begin{aligned} \mathcal{X}_j &\subset \{x \mid y(x) = j\} \text{ and } |\mathcal{X}_j| = G \text{ and} \\ \mathcal{Z}_j &= \{\{x\} \mid x \in \mathcal{I} \setminus \mathcal{X}_j \text{ and } y(x) = j\} \end{aligned}$$

A CRF model for identity inference I

CRF definition

CRF defined by a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, a set of random variables $\mathcal{Y} = \{Y_j \mid j = 1 \dots |V|\}$ and a set of possible labels $\mathcal{L}$.

- Vertices $\mathcal{V}$: index the random variables in $\mathcal{Y}$
- Edges $\mathcal{E}$: statistical dependence relations between random variables

A CRF model for identity inference I

CRF definition

CRF defined by a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, a set of random variables $\mathcal{Y} = \{Y_j \mid j = 1 \dots |V|\}$ and a set of possible labels $\mathcal{L}$.

- Vertices $\mathcal{V}$: index the random variables in $\mathcal{Y}$
- Edges $\mathcal{E}$: statistical dependence relations between random variables

The labeling problem

Find an assignment of labels to nodes that minimizes an energy function E over possible labelings $\mathbf{y}^* = (y_i^*)_{i=1}^{|V|}$:

$$\tilde{\mathbf{y}} = \arg\min_{\mathbf{y}^*} E(\mathbf{y}^*) \tag{1}$$

A CRF model for identity inference II

Energy function definition

$$E(\mathbf{y}^*) = \sum_{i \in \mathcal{V}} \phi_i(y_i^*) + \lambda \sum_{(i,j) \in \mathcal{E}} \psi_{ij}(y_i^*, y_j^*), \tag{2}$$

- $\phi_i(y_i^*)$ unary data potential, cost of assigning label y_i^* to vertex i
- $\psi_{ij}(y_i^*, y_j^*)$ binary smoothness potential, conditional cost of assigning labels y_i^* and y_j^* respectively to vertices i and j
- Parameter λ: tradeoff between data and smoothness costs.

Efficient algorithms for finding the optimal labeling $\tilde{\mathbf{y}}$, for example, **graph cuts** [Kolmogorov and Zabin, 2004, Szeliski et al., 2008].

Definition of unary data and smoothness potentials

Unary data potential

Cost of assigning label y_i^* to vertex i given $\mathbf{x}(x_i)$. Proportional to the minimum distance between x_i and any gallery image of individual y_i^*.

$$\phi_i(y_i^*) = \begin{cases} 1 & \text{if } x_i \in \mathcal{X} \text{ and } y_i^* \neq y(x_i) \\ \min_{x \in \mathcal{X}_{y_i^*}} ||\mathbf{x}(x) - \mathbf{x}(x_i)|| & \text{otherwise.} \end{cases}$$

Definition of unary data and smoothness potentials

Unary data potential

Cost of assigning label y_i^* to vertex i given $\mathbf{x}(x_i)$. Proportional to the minimum distance between x_i and any gallery image of individual y_i^*.

$$\phi_i(y_i^*) = \begin{cases} 1 & \text{if } x_i \in \mathcal{X} \text{ and } y_i^* \neq y(x_i) \\ \min_{x \in \mathcal{X}_{y_i^*}} ||\mathbf{x}(x) - \mathbf{x}(x_i)|| & \text{otherwise.} \end{cases}$$

Smoothness potential

Encourage similar detections to share the same labels:

$$\psi_{ij}(y_i^*, y_j^*) = w_{ij} \min_{\substack{x \in \mathcal{X}_{y_i^*} \\ x' \in \mathcal{X}_{y_j^*}}} ||\mathbf{x}(x) - \mathbf{x}(x')||. \quad (3)$$

Weighting factors w_{ij} allow flexibility in the smoothness potential between nodes i and j.

Solving re-identification problem through the CRF

An identity inference problem $\mathcal{R} = (\mathcal{X}, \mathcal{Z})$ is mapped onto a CRF by defining the vertex $\mathcal{V}$ and edge $\mathcal{E}$ sets in terms of the gallery $\mathcal{X}$ and test $\mathcal{Z}$ images sets.

MvsM re-identification problem

$$\mathcal{V} = \bigcup_{i=1}^{M} \mathcal{Z}_i \text{ and } \mathcal{E} = \{(x_i, x_j) \mid x_i, x_j \in \mathcal{Z}_l \text{ for some } l\}.$$

The edge topology in this CRF is completely determined by the group structure as expressed by the $\mathcal{Z}_j$ ($w_{ij}=1$).

Solving identity inference problem through the CRF

General identity inference case (as well as in SvsAll re-identification) no identity grouping information is available for the test set

Identity inference problem

$$\mathcal{V} = I \text{ and } \mathcal{E} = \bigcup_{x_i \in \mathcal{V}} \{(x_i, x_j) \mid x_j \in \mathrm{kNN}(x_i)\},$$

where the $\mathrm{kNN}(x_i)$ maps an image to its k most similar images in feature space.

Topology of this CRF formulation uses feature similarity to form connections between nodes.
Edges defined as K ($K = 4$) nearest neighbors in feature space.

Solving identity inference problem through the CRF

Smoothness potential weighting

- Weights w_{ij} of (3) between vertices i and j include feature similarity and (eventually) a temporal constraint:

$$w_{ij} = (1-\alpha)(1-||\mathbf{x}(x_i)-\mathbf{x}(x_j)||)+\alpha\tau_{ij}, \tag{4}$$

$\alpha \in [0,1]$ tradeoff between temporal and feature similarities (α=0.3)

- τ_{ij} is a temporal weighting factor:

$$\tau_{ij} = \begin{cases} 1-\frac{|f_i-f_j|}{\tau} & \text{if } |f_i-f_j| \leq \tau \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

where f_i and f_j are the frame numbers in which detections i and j occurred, respectively, and τ is a threshold limiting the temporal influence to a finite number of frames (fixed to 25).

Experimental protocol

Publicly available ETHZ [Schwartz and Davis, 2009] dataset: 3 video sequences (ETHZ1: 4857 images, 83 persons; ETHZ2: 1961 images, 35 persons; ETHZ3: 1762 images, 28 persons).
On average each person appears in more than 50 images.

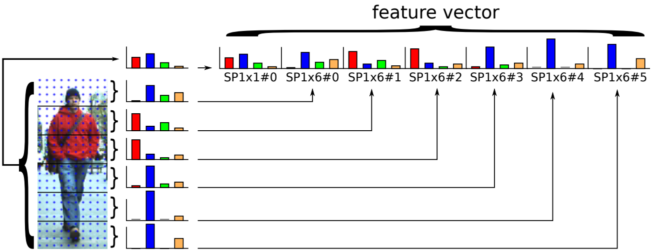

Figure 1: Our feature descriptor: 1×6 spatial pyramid histogram over densely sampled HueSIFT [van de Weijer and Schmid, 2006] features quantized to a visual vocabulary of 512 visual words.

Re-identification

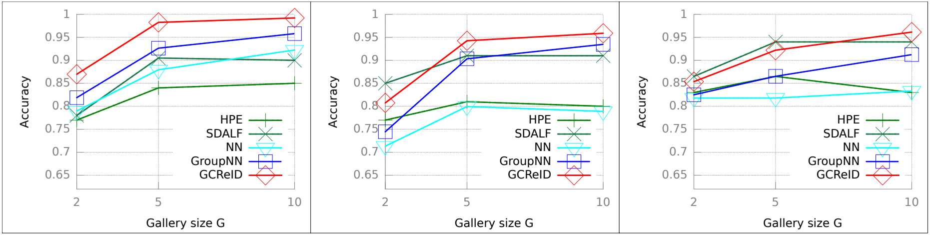

Figure 2: MvsM re-identification accuracy ($\lambda = 1$). Left to right: ETHZ1, ETHZ2 and ETHZ3. These are *not* CMC curves, but are Rank-1 *classification* accuracies over varying gallery and test set sizes.

CRF enforces labeling consistency, allows our approach to outperform simpler GroupNN and state-of-the-art methods.

Identity inference

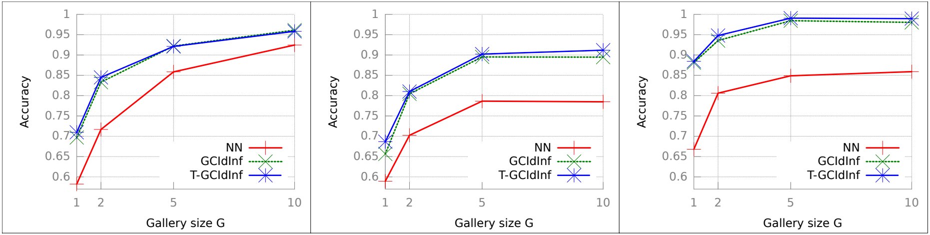

Figure 3: Identity inference accuracy on ETHZ datasets ($\lambda = 5$). Left to right: ETHZ1, ETHZ2 and ETHZ3.

- CRF framework clearly improves accuracy over simple NN model.
- T-GCIDInf: feature similarity-weighted edges with temporal constraints (avg improvement of 15% over NN)
- Label many unknown images using only few gallery images
- Robustness to occlusions and illumination changes

Qualitative results

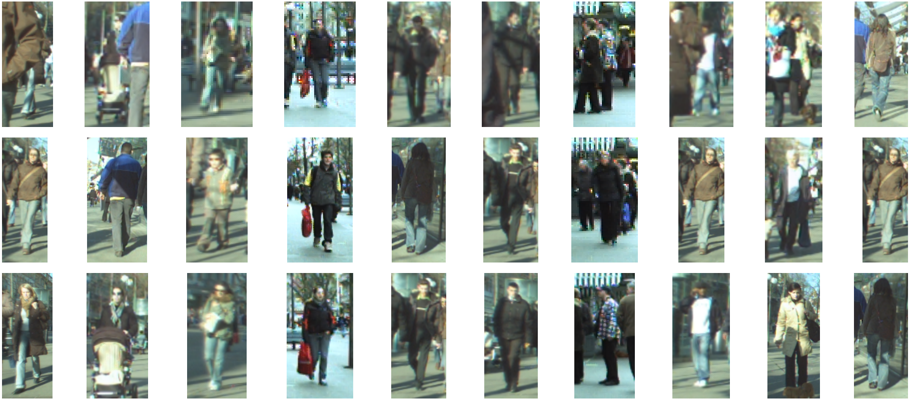

Figure 4: Identity inference results (SvsAll). First row: test image, second row: incorrect NN result, third row: correct result given by GCIdInf.

Neighbors graph visualization

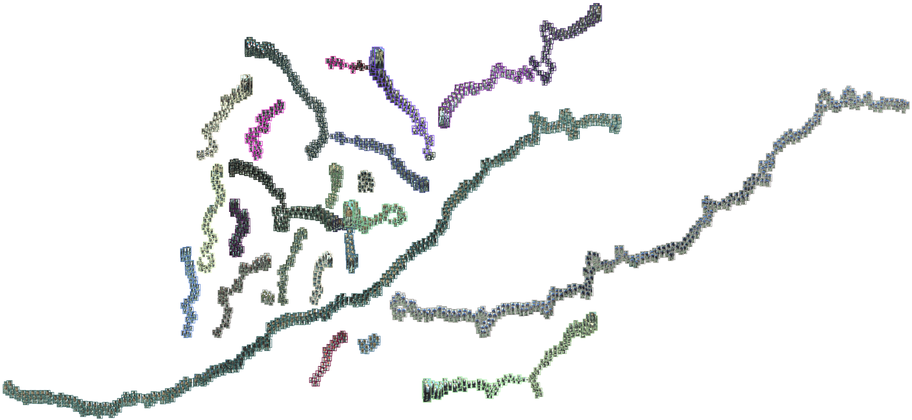

Figure 5: ETHZ3 neighbors graph visualization

Discussion

Conclusion

- Identity inference, generalization of re-identification scenarios:
  - Generalization of the single-versus-all scenario
  - Relaxation of the multi-versus-multi shot case
  - No hard knowledge about relationships between test images (e.g. that they correspond to the same individual) required.
- Solved by CRF-based approach. Neighborhood topology defined by feature space and temporal (when available) similarity
- Experimental results:
  - CRF approach can efficiently solve standard re-identification tasks
  - CRF model can also solve more general identity inference problems

Future works on exploring more powerful descriptors and more realistic configurations for identity inference in the real world. Recording of a multi-camera dataset for identity inference.

References

S. Bak, E. Corvee, F. Bremond, and M. Thonnat. Multiple-shot human re-identification by mean riemannian covariance grid. In *Proceedings of AVSS*, pages 179–184, 2011.

Loris Bazzani, Marco Cristani, Alessandro Perina, and Vittorio Murino. Multiple-shot person re-identification by chromatic and epitomic analyses. *Pattern Recognition Letters*, 33(7):898–903, 2012.

Y. Cai and M. Pietikäinen. Person re-identification based on global color context. In *Proceedings of ACCV Workshops*, pages 205–215, 2011.

Dong S. Cheng, Marco Cristani, Michele Stoppa, Loris Bazzani, and Vittorio Murino. Custom pictorial structures for re-identification. In *Procedings of BMVC*, 2011.

M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. *Proceedings of CVPR*, pages 2360–2367, 2010.

Douglas Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *Proceedings of ECCV*, pages 262–275, 2008. doi: http://dx.doi.org/10.1007/978-3-540-88682-2_21.

V. Kolmogorov and R. Zabin. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):147–159, 2004.

Bryan Prosser, Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Person re-identification by support vector ranking. In *Proceedings of BMVC*, 2010.

W.R. Schwartz and L.S. Davis. Learning discriminative appearance-based models using partial least squares. In *Proceedings of SIBGRAPI*, pages 322–329. IEEE, 2009.

R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother. A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(6):1068–1080, 2008.

Joost van de Weijer and Cordelia Schmid. Coloring local feature extraction. In *Proceedings of ECCV*, volume Part II, pages 334–348, 2006. URL `http://lear.inrialpes.fr/pubs/2006/VS06`.

B. Yang and R. Nevatia. An online learned crf model for multi-target tracking. In *Proceedings of CVPR*, 2012.

Questions?

Identity inference: generalizing person re-identification scenarios

Svebor Karaman Andrew D. Bagdanov

MICC - Media Integration and Communication Center
University of Florence
svebor.karaman@unifi.it, bagdanov@dsi.unifi.it

1st International Workshop on Re-Identification
12 October 2012