Transitive Re-identification

Yulia Brand[1]
yulias@techunix.technion.ac.il

Tamar Avraham[2]
http://www.cs.technion.ac.il/~tammya

Michael Lindenbaum[2]
http://www.cs.technion.ac.il/~mic

[1] Electrical Engineering Department

[2] Computer Science Department

Technion - I.I.T.
Haifa 32000, Israel

Person re-identification (ReID) problem has lately received increasing attention especially due to its important role in surveillance systems, which should be able to keep track of people after they have left the field of view of one camera and entered the field of view of any overlapping or non-overlapping camera.

ReID accuracy can be significantly improved given a training set that demonstrates changes in appearances associated with the two non-overlapping cameras involved. Here we test whether this advantage can be maintained when directly annotated training sets are not available for all camera-pairs at the site. Given the training sets capturing correspondences between cameras A and B and a different training set capturing correspondences between cameras B and C, the Transitive Re-IDentification algorithm (TRID) suggested here provides a classifier for (A,C) appearance pairs; see Fig. 1.

TRID establishes a path between the non-directly trainable camera pair (A,C) by marginalization over the domain of possible appearances in camera B. Camera B plays the role of the 'connecting element' between cameras A and C. Our goal is to estimate the conditional probability $P(Y_{AC}|x_A,x_C)$ where the notations x_A and x_C refer to a feature vector describing the appearance observed by cameras A and C and the binary variable Y_{AC} gets the value 1 if and only if the appearances given in A and in C are of the same identity. When the feature vector is known to correspond to a particular individual of identity i, we denote it by x_A^i. Thus, the pair $\{(x_A^i, x_C^i)\}$ is a pair of feature vectors corresponding to the same person but to different cameras. Although a training set consisting of annotated pairs $\{(x_A^i, x_C^i)\}$ is not available, we can exploit the annotated sets $S_{AB} = \{(x_A^i, x_B^i)\}$, $i = 1,...,n$ and $S_{BC} = \{(x_B^j, x_C^j)\}$, $j = n+1,...,n+m$. The desired conditional probability is expressed by:

$$P(Y_{AC}|x_A,x_C) = \frac{\int_{x_B} P(Y_{AB}|x_A,x_B)P(Y_{BC}|x_B,x_C)f_{X_B}(x_B)dx_B}{1 - \int_{x_B} P(\overline{Y_{AB}}|x_A,x_B)P(\overline{Y_{BC}}|x_B,x_C)f_{X_B}(x_B)dx_B} \quad . \tag{1}$$

In our experiments we found that the method is superior to that of a state-of-the-art camera invariant method, which used a fixed similarity measure (SDALF [2]). We compared the transitive TRID performance to direct learning from (A,C) examples using the ICT algorithm [1], which has shown state-of-the-art results for modeling the transfer of appearances associated with two specific cameras. We found that the TRID is also more accurate than a method that makes a naive use of training sets associated with (A,B) and (B,C) camera-pairs by just applying the ICT algorithm trained by the union of S_{AB} and S_{BC}; see Fig. 2. In certain

(a) [A B C]=[3 5 7] (b) [A B C]=[1 3 8]
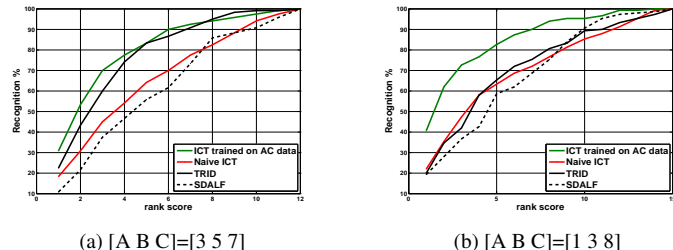
Figure 2: CMC curves comparing the performance of TRID to that of other, non-transitive algorithms, over 2 combinations of camera-triplets from the SAIVT-SoftBio dataset [3] (see more results in the paper).

(a) (b) (c)

Figure 3: Illustration of a system consisting of N cameras $C_1, C_2, ..., C_N$: (a) Inter-camera training sets are available only for $N-1$ *directly trainable pairs* (C_i, C_{i+1}), $i = 1,...,N-1$. (b) The transitive algorithm infers classifiers for the *non-directly trainable pairs* (C_i, C_{i+2}). (c) Inter-camera classifiers for any other camera pair in the system can be deduced by recursively applying the transitive algorithm.

conditions TRID even meets the performance achieved when direct (A,C) training data is available.

Given a camera network of N cameras, and *directly trainable pairs* (C_i, C_{i+1}), $i = 1,...,N-1$ (see Fig. 3(a)), the proposed algorithm enables the inference of a classifier for the *non-directly trainable pair* (C_i, C_{i+2}); see Fig. 3(b). The inter-camera classifiers for any other camera pair can be deduced by recursively applying the transitive algorithm; see Fig. 3(c). Moreover, when adding camera $(N+1)$, only one inter-camera training set is required for establishing all correspondences. This approach significantly reduces the annotation effort inherent in a learning system, which goes down from $O(N^2)$ to $O(N)$.

To the best of our knowledge, this work is not only the first approach to transitivity in ReID but also, more generally, transitivity in domain adaptation. In our future work we intend to test the TRID algorithm on other domain adaptation tasks, and to examine alternative transitive approaches.

[1] T. Avraham, I. Gurvich, M. Lindenbaum, and S. Markovitch. Learning implicit transfer for person re-identification. In *The 1st International Workshop on Re-Identification (Re-Id 2012), in conjunction with ECCV, LNCS*, pages 381–390, 2012.

[2] L. Bazzani, M. Cristani, and V. Murino. Symmetry-driven accumulation of local features for human characterization and re-identification. *Computer Vision and Image Understanding*, 117:130–144, 2013.

[3] A. Bialkowski, S. Denman, P. Lucey, S. Sridharan, and C. B. Fookes. A database for person re-identification in multi-camera surveillance networks. In *Digital Image Computing: Techniques and Applications (DICTA 2012)*, pages 1–8, 2012.
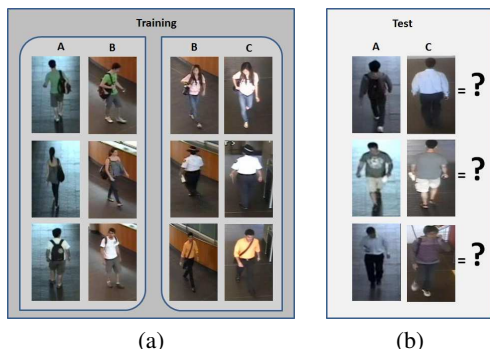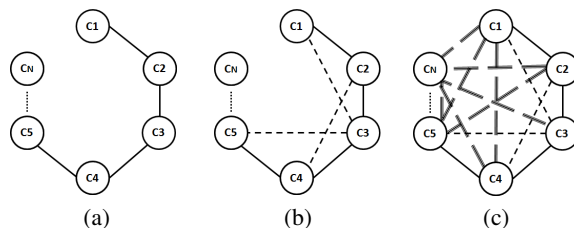
(a) (b)

Figure 1: The transitive ReID setup. Given the training sets S_{AB} ((a)left) capturing correspondences between people $1...n$ and a training set S_{BC} ((a)right) capturing correspondences between different people $n+1,...,m+n$, we would like to be able to classify (A,C) pairs (b) without an S_{AC} training set.