Transitive Re-identification

Yulia Brand[1] Tamar Avraham[2] Michael Lindenbaum[2]

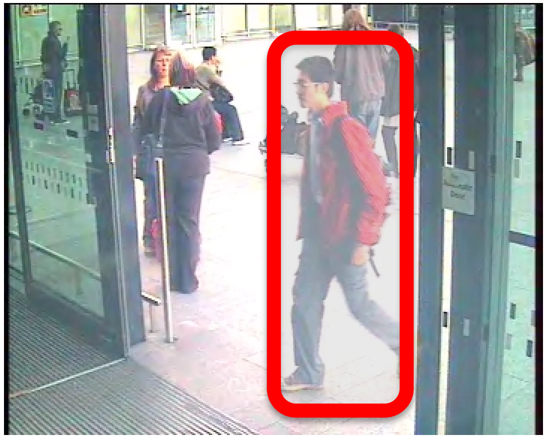
[1]Electrical Engineering Department
[2]Computer Science Department
Technion - I.I.T. Haifa, Israel

This research was supported by the MAGNET program in the Israeli ministry of industry and commerce, by the Israeli ministry of science and by the E. and J. Bishop research fund.

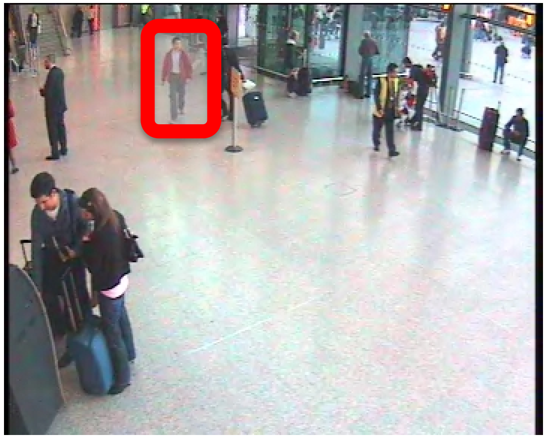
BMVC 2013

ReIDentification (ReID)

recognizing individuals over different camera views
(we focus on ReID based on appearance based cues)



Camera A, time t



Camera B, time t+Δt

- **upper image from:** A. Bialkowski, S. Denman, P. Lucey, S. Sridharan, and C. B. Fookes. "A database for person re-identification in multi-camera surveillance networks." (DICTA 2012)
- **lower images:** courtesy of Marco Cristani

ReIDentification

Appearance-based ReIDentification methods:

Learning based methods:

- require a training set
- classification is with a classifier or a metric learned on the training set
- ***camera-invariant*** or ***camera-specific***, depending on the used training set

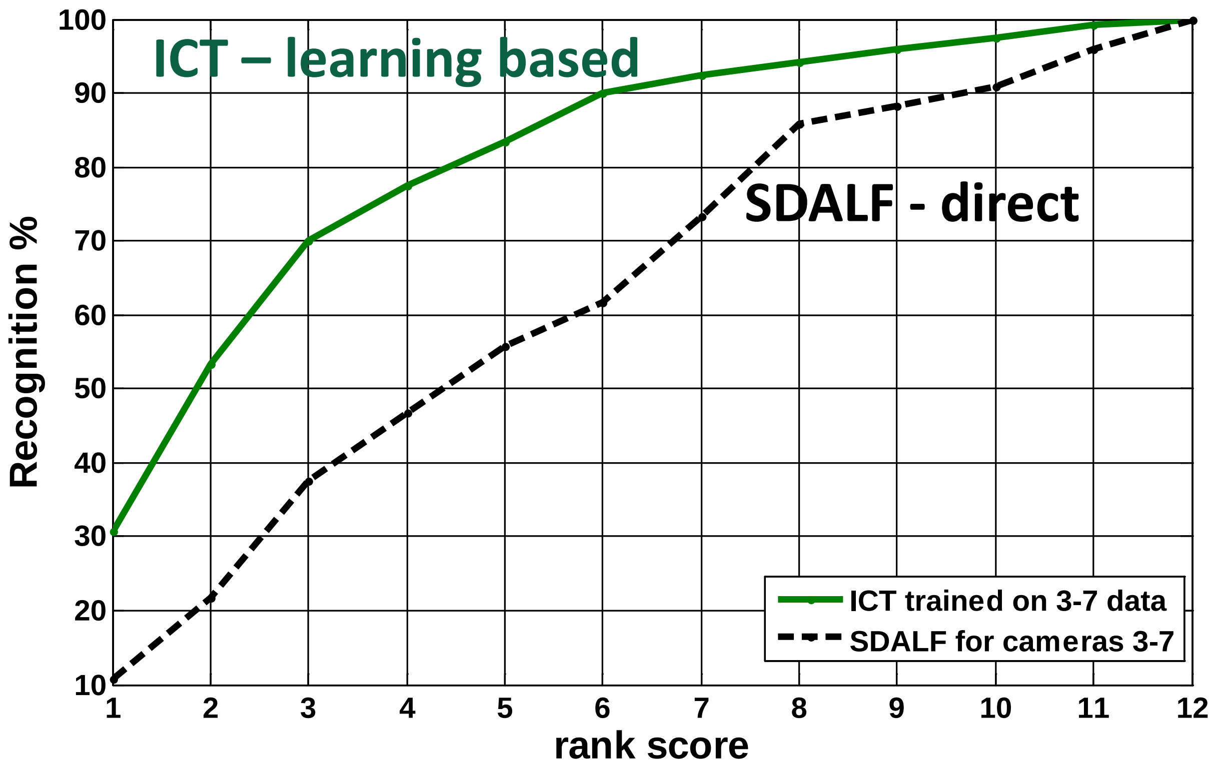
Direct methods:

- do not depend on any training set
- classification is based on pre-defined distance metric
- always ***camera-invariant***

perform better for "camera-specific" scenarios:
ReID benefits from training with corresponding appearance-pairs captured by specific camera pairs, as the background, illumination, resolution and pose are camera dependent.

ReIDentification

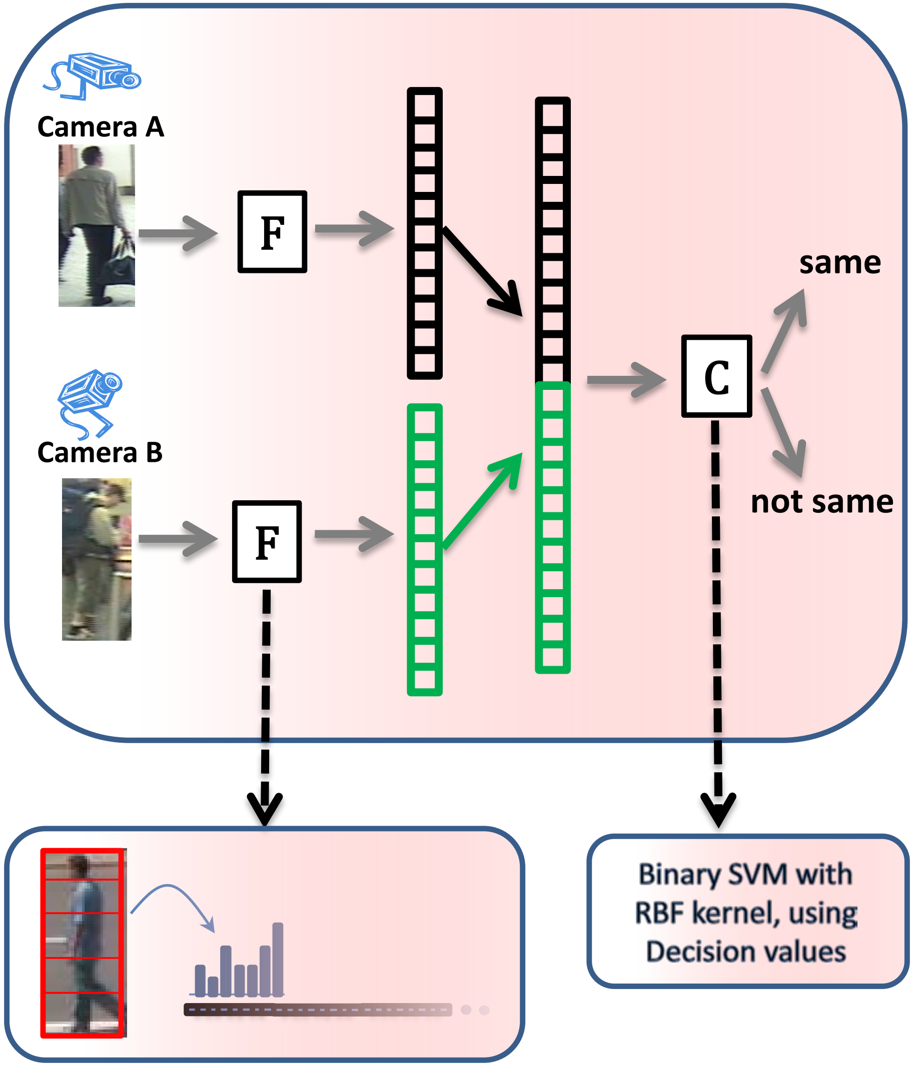
Learning based methods perform better for “camera-specific” scenarios:



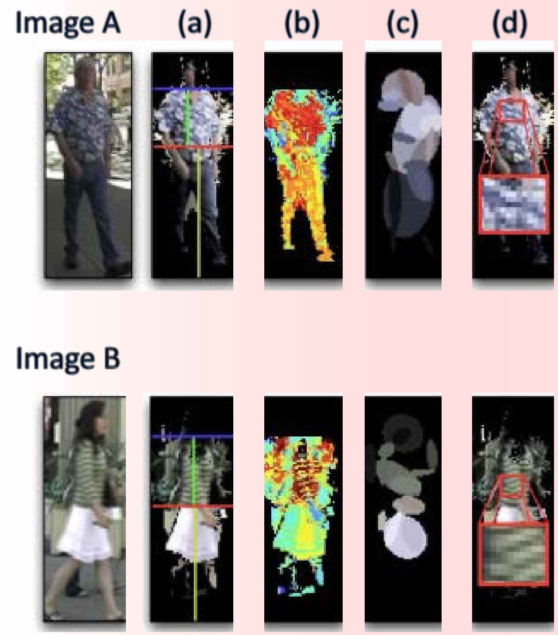
- ICT (Implicit Camera Transfer) by Avraham et al (ECCV Re-Id 2012)
- SDALF (Symmetry-Driven Accumulation of Local Features) by Farenzena et al (CVPR 2010) and Bazzani et al (CVIU 2013)
- performance tested on 2 cameras from the SAIVT-SoftBio dataset

ReIDentification

Learning method: ICT



Direct method: SDALF



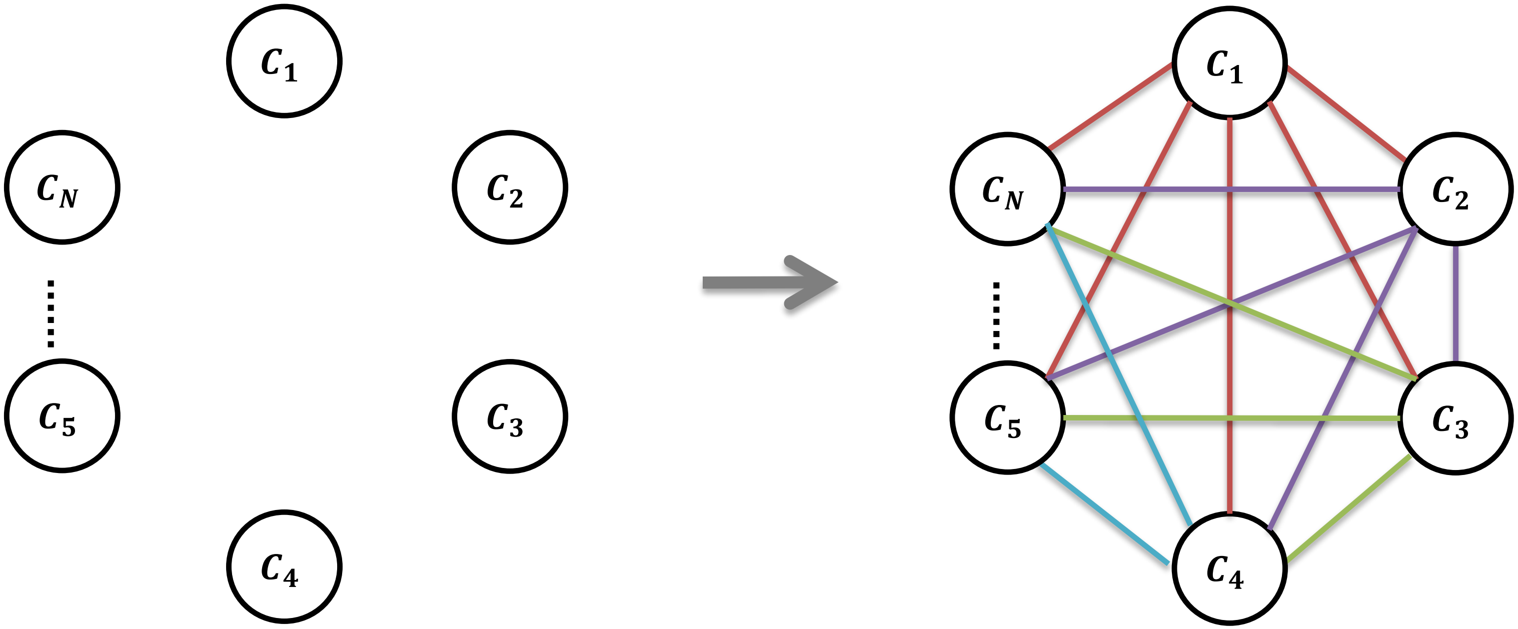
(a) localizing meaningful body parts
(b) weighted HSV histogram
(c) MSCR - maximally stable color regions
(d) RHSP - recurrent highly structured patches
=> matching SDALF descriptors

* Image from: L. Bazzani, M. Cristani, and V. Murino. Symmetry-driven accumulation of local features for human characterization and re-identification.

Motivation

Given a site with N cameras $C_1, C_2, \ldots C_N$, we may ReID either by:

1. applying a direct method
2. applying a learning based method where distinct inter-camera training sets must be collected and annotated for each camera-pair. This is impractical.

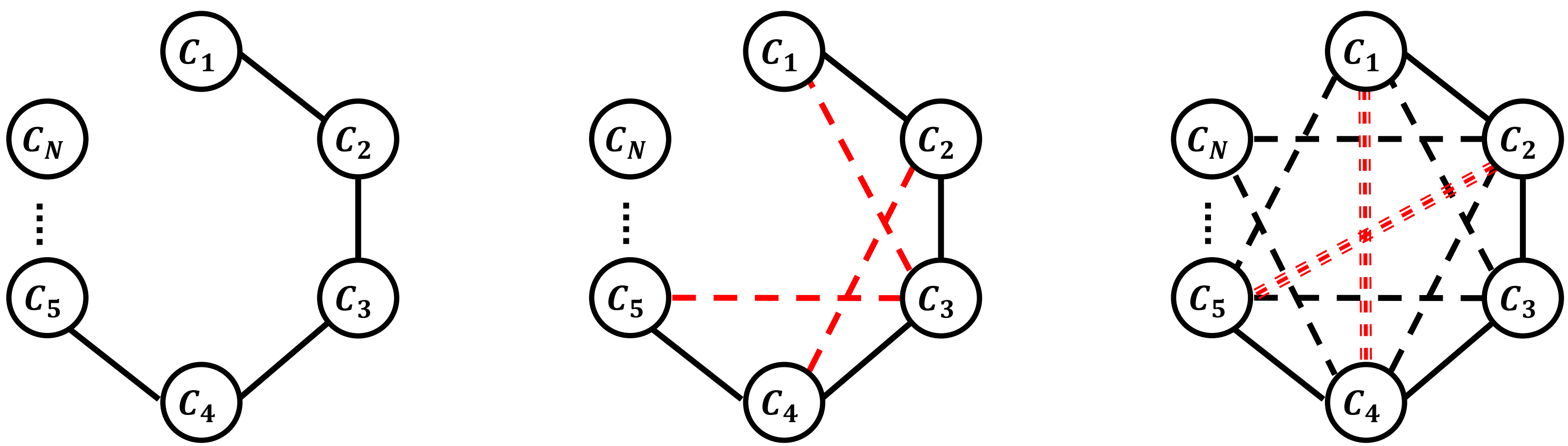


Motivation

We aim at reducing the number of required direct inter-camera training sets from $O(N^2)$ to $O(N)$ by suggesting a **transitive algorithm** which uses inter-camera training sets only for $N-1$ camera-pairs $(C_i, C_{i+1}), i = 1, \ldots N-1$, and infers a ReID classifier for any other camera-pair in the system.

By applying the transitive algorithm:

By recursively applying the transitive algorithm:

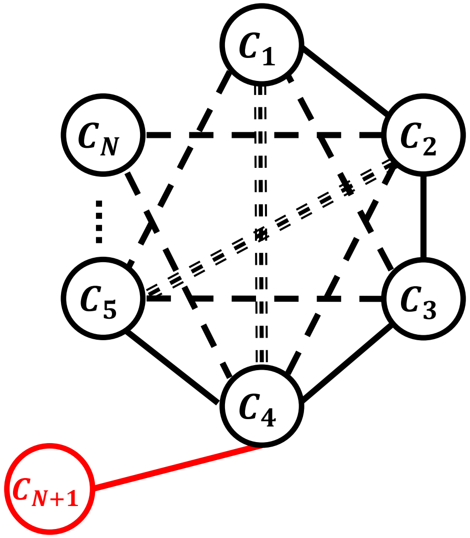


directly trainable pairs

non-directly trainable pairs

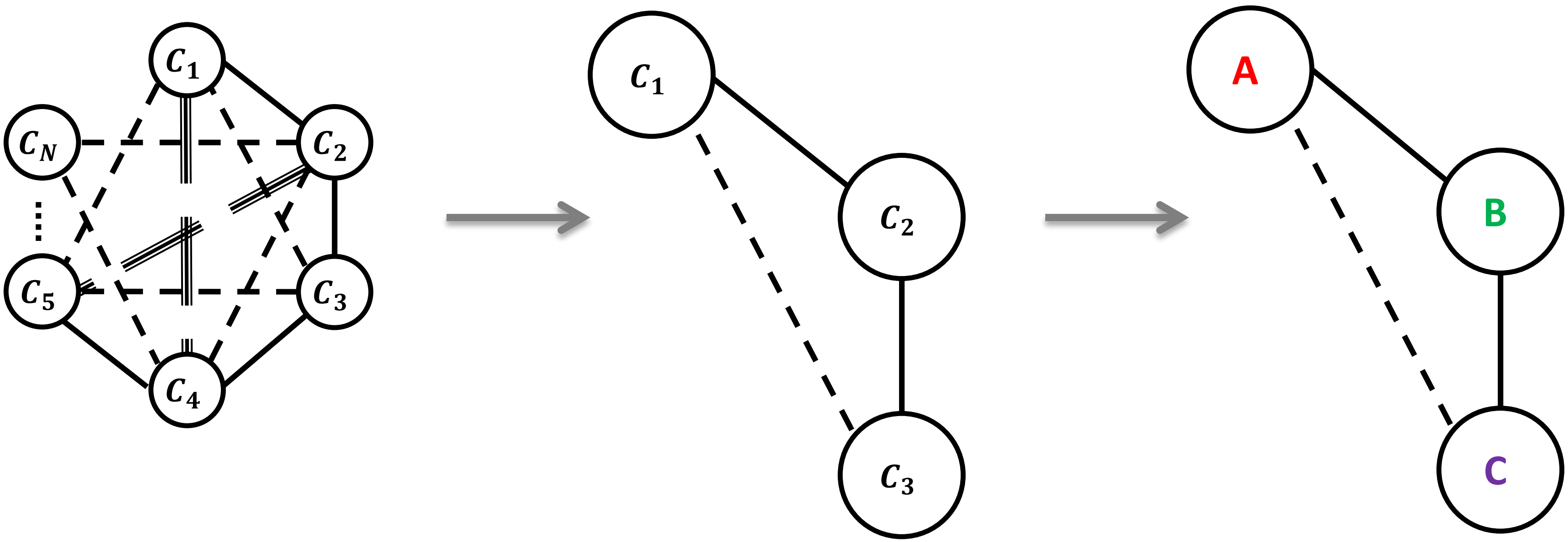
Motivation

When a new camera is added to a set of N previously installed cameras, with the proposed transitive approach, collecting data for only one pair is required.

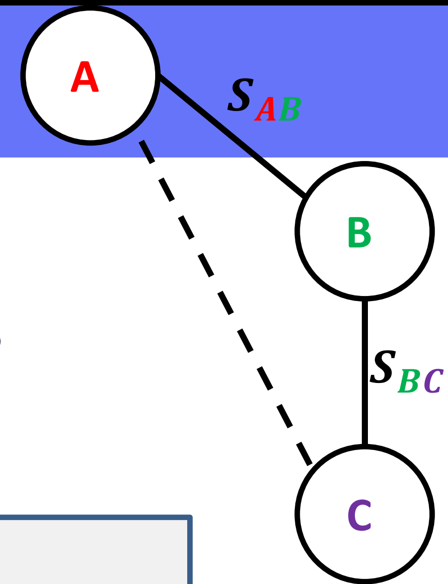


The Transitive Algorithm

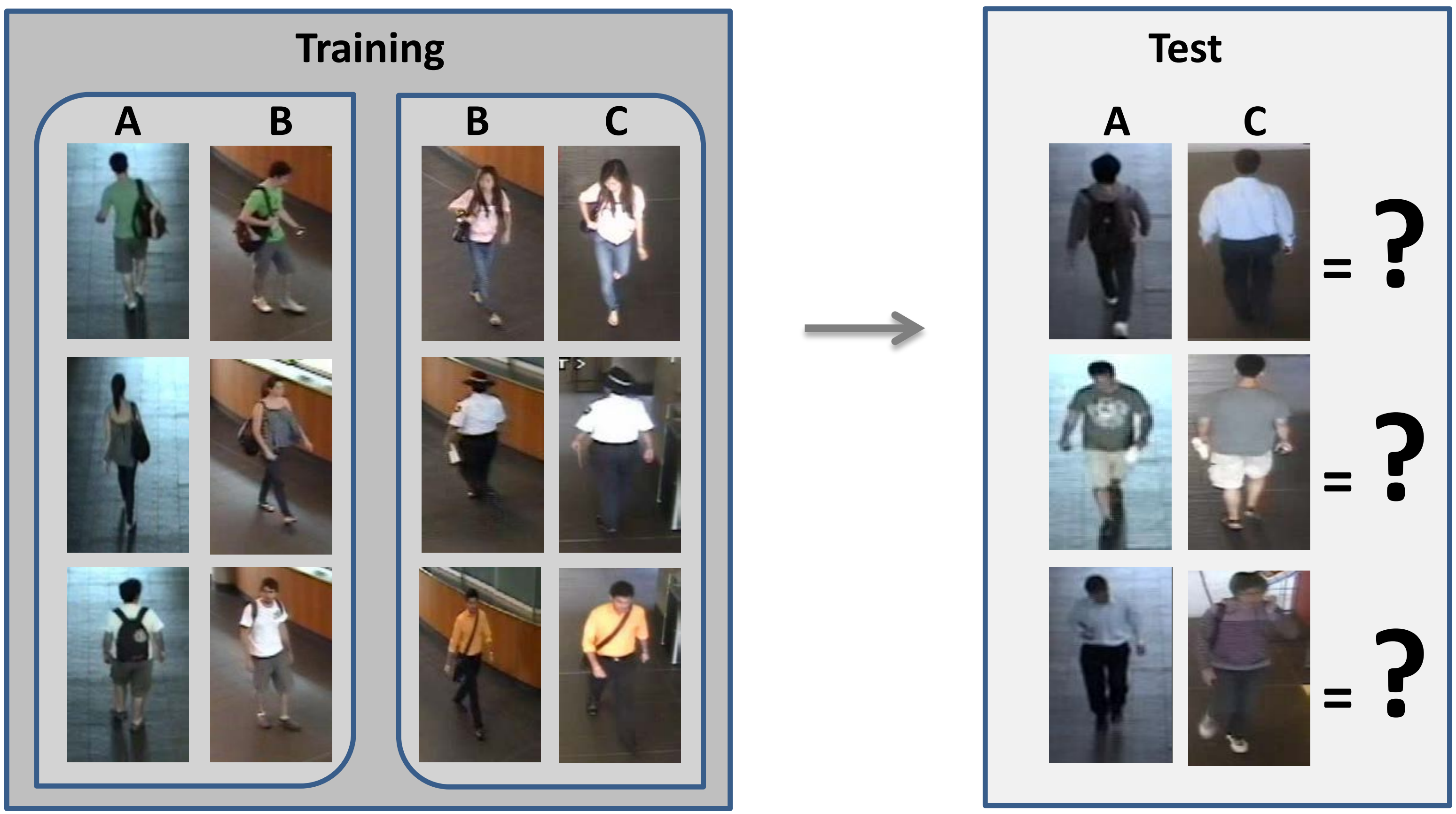
- We focus on a camera triplet case $[A\ B\ C]$.



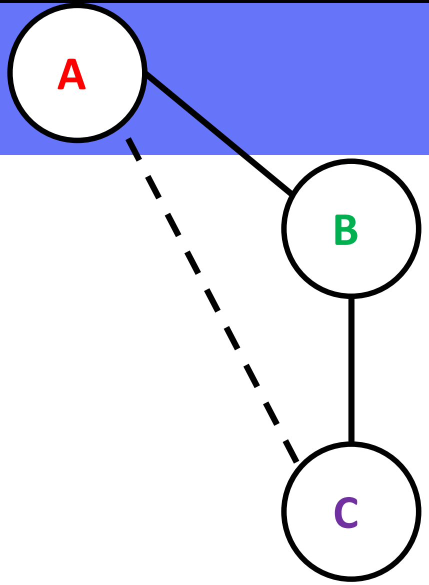
The Transitive Algorithm



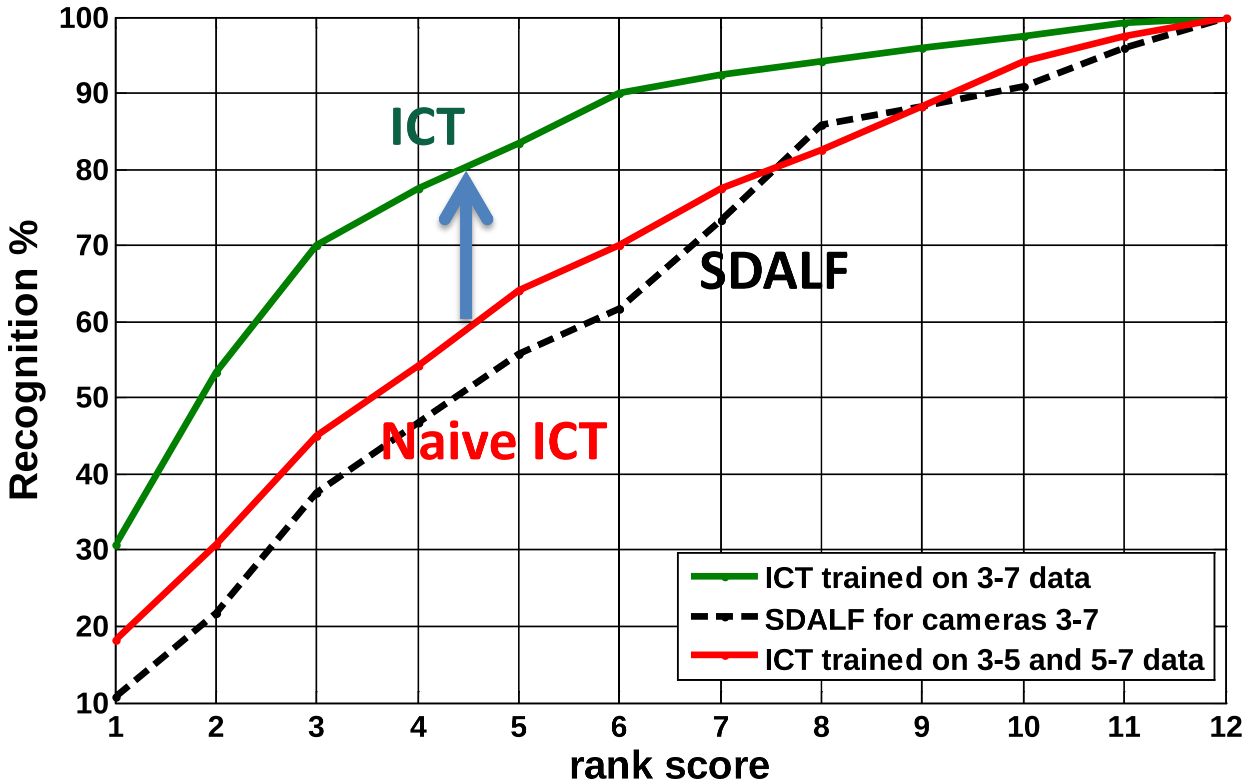
- Given two training sets S_{AB}, S_{BC} associated with camera-pairs (A, B) and (B, C) , respectively, we would like to infer a classifier for the (A, C) camera-pair, for which S_{AC} is missing.



The Transitive Algorithm

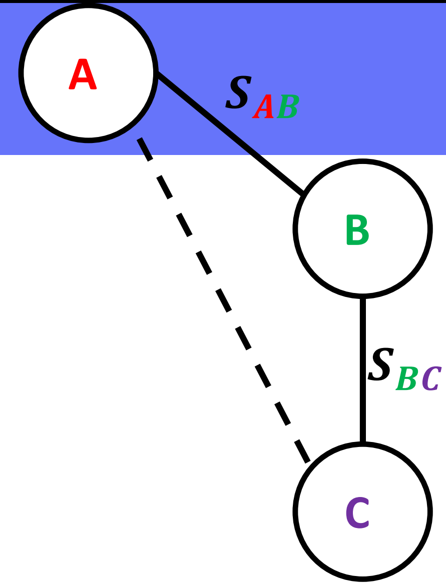


Expanding the learning based method – the Naive solution: training the (A, C) classifier with $S_{AB} \cup S_{BC}$



Our goal is to find a more effective use of the available training sets that will decrease this performance gap.

The Transitive Algorithm



$$S_{AB} = \{(x_A^i, x_B^j)\}$$

$$S_{BC} = \{(x_B^i, x_C^j)\}$$

$$Y_{AB}(x_A^i, x_B^j) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

$$Y_{BC}(x_B^i, x_C^j) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

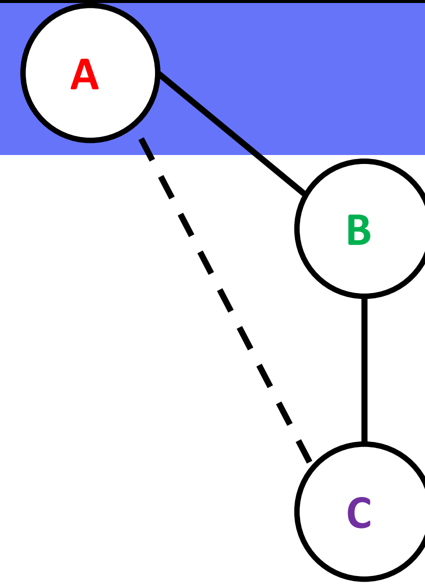
$$\boxed{P(Y_{AC}|x_A, x_C) = ?}$$

The Transitive ReID algorithm (TRID) establishes a path between the non-directly trainable camera pair (A, C) by **marginalization** over a **"connecting element"**, the domain of possible appearances in camera B.

$$\mathrm{P}(Y_{AC}|x_A, x_C) = \sum_{y_{AB}\in\{0,1\}} \sum_{y_{BC}\in\{0,1\}} \int_{x_B\in R^d} P(Y_{AC}, Y_{AB} = y_{AB}, Y_{BC} = y_{BC}, x_B \,|\, x_A, x_C)\, dx_B$$

$$\mathrm{P}(Y_{AC}|x_A, x_C) = \sum_{y_{AB}\in\{0,1\}} \sum_{y_{BC}\in\{0,1\}} \left[\int_{x_B\in R^d} P(Y_{AC}, Y_{AB} = y_{AB}, Y_{BC} = y_{BC} \,|\, x_A, x_B, x_C) f_{x_B}(x_B)\, dx_B \right]$$

The Transitive Algorithm



$$S_{AB} = \{(x_A^i, x_B^j)\}$$

$$S_{BC} = \{(x_B^i, x_C^j)\}$$

$$Y_{AB}(x_A^i, x_B^j) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

$$Y_{BC}(x_B^i, x_C^j) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

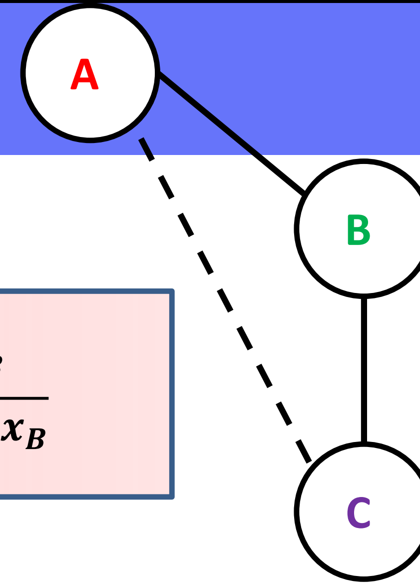
$$\boxed{P(Y_{AC} | x_A, x_C) = ?}$$

The Transitive ReID algorithm (TRID) establishes a path between the non-directly trainable camera pair (A, C) by **marginalization** over a **"connecting element"**, the domain of possible appearances in camera B.

$\vdots$

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

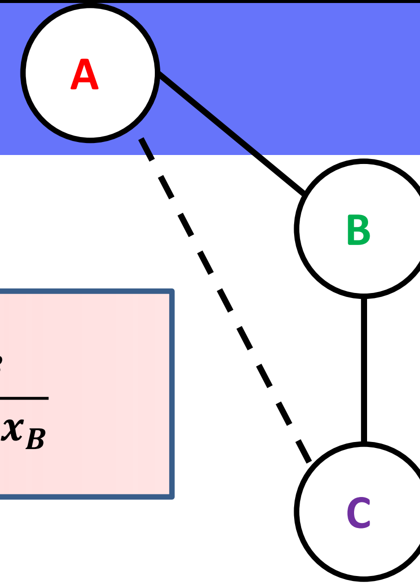
The Transitive Algorithm



$$P(Y_{AC}|x_A, x_C) = \frac{\int_{x_B} \underbrace{P(Y_{AB}|x_A, x_B)}_{\text{ICT [A B]}} \underbrace{P(Y_{BC}|x_B, x_C)}_{\text{ICT [B C]}} f_{X_B}(x_B)\, dx_B}{1 - \int_{x_B} P(\overline{Y_{AB}}|x_A, x_B) P(\overline{Y_{BC}}|x_B, x_C)\, f_{X_B}(x_B)\, dx_B}$$

- $P(Y_{AB}|x_A, x_B)$ - probability for a match in camera-pair (A, B).
- Any classifier can be used here provided that it can be modified to output probability.
- TRID uses the ICT algorithm by training two ICT ReID classifiers. The SVM decision values are converted to probability estimates using a sigmoid according to Platt's method.

The Transitive Algorithm

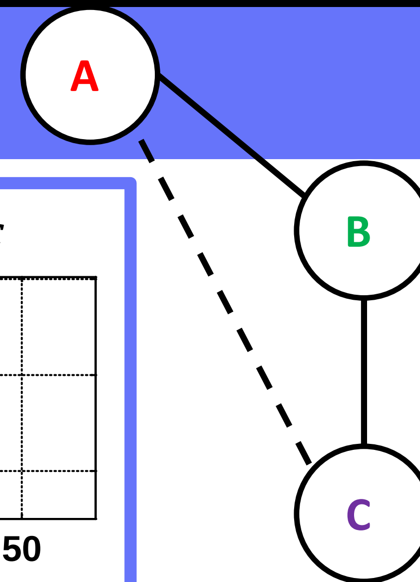


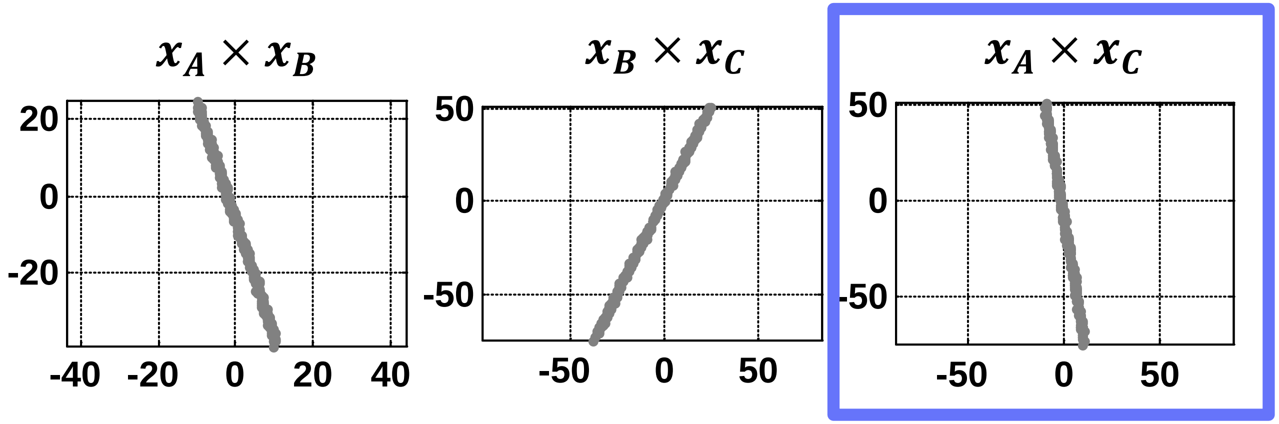
$$P(Y_{AC}|x_A,x_C) = \frac{\boxed{\int_{x_B}} P(Y_{AB}|x_A,x_B)P(Y_{BC}|x_B,x_C)\,\boxed{f_{X_B}(x_B)}\,dx_B}{1-\int_{x_B} P(\overline{Y_{AB}}|x_A,x_B)P(\overline{Y_{BC}}|x_B,x_C)\,f_{X_B}(x_B)\,dx_B}$$

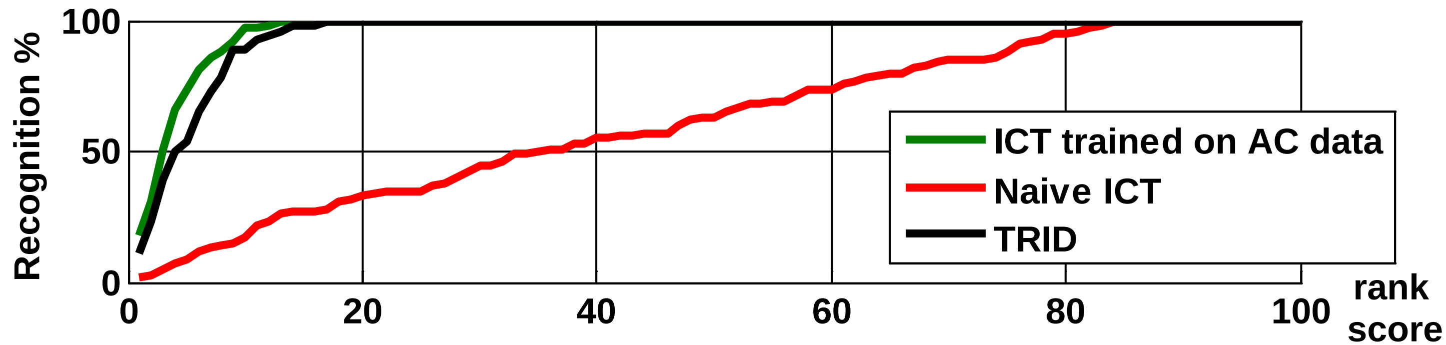
- $f_{X_B}(x_B)$ is a multi-dimensional probability density function.
- Can be estimated by methods for density estimation using all $x_B \in S_B$.
- However, estimating high-dimensional density is hard and the integration is computationally costly.
- Thus, in TRID the integral is approximated by a sum relying on the smoothness of the probability functions.

$$P(Y_{AC}|x_A,x_C) \approx \frac{\frac{1}{|S_B|}\sum_{x_B\in S_B} P(Y_{AB}|x_A,x_B)P(Y_{BC}|x_B,x_C)}{1-\frac{1}{|S_B|}\sum_{x_B\in S_B} P(\overline{Y_{AB}}|x_A,x_B)P(\overline{Y_{BC}}|x_B,x_C)}$$

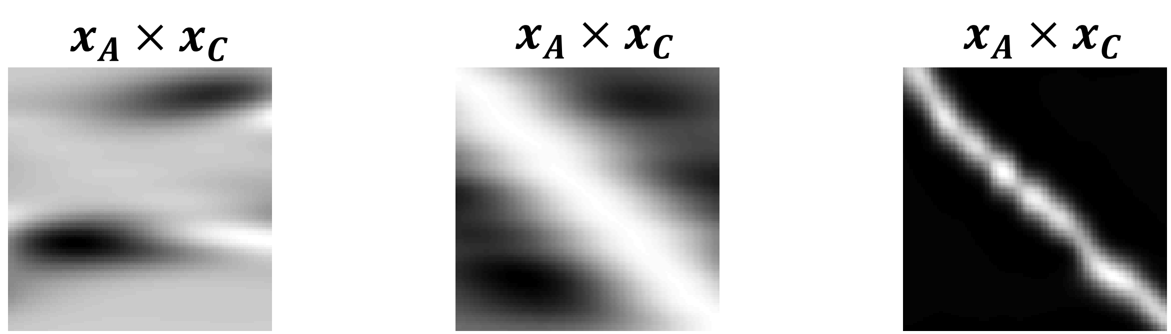
Synthetic Experiment 1





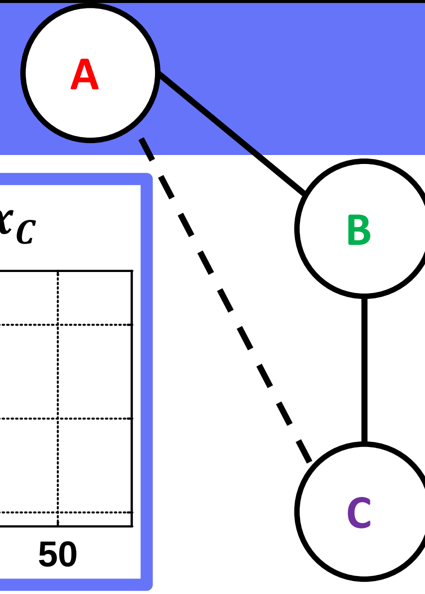


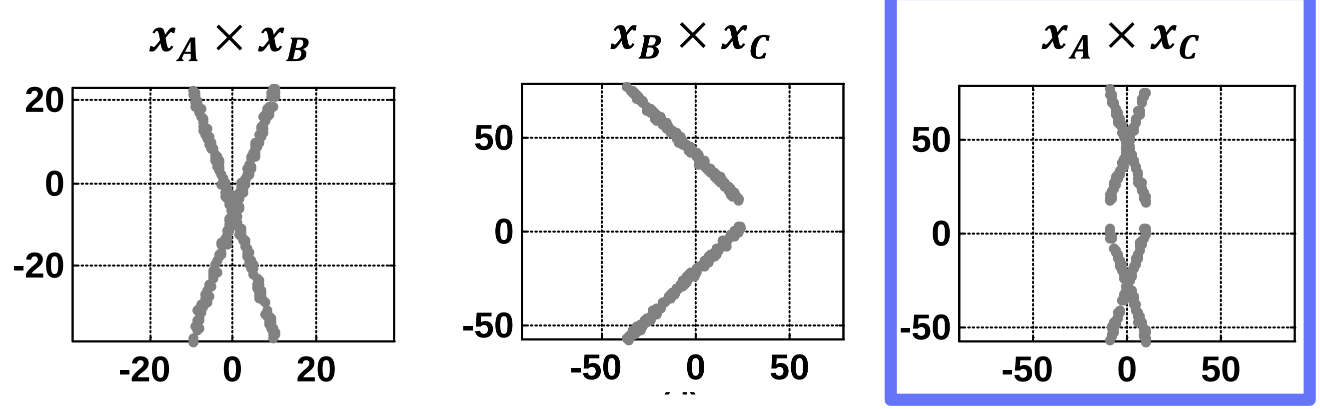
$P(Y_{AC}|x_A, x_C)$:

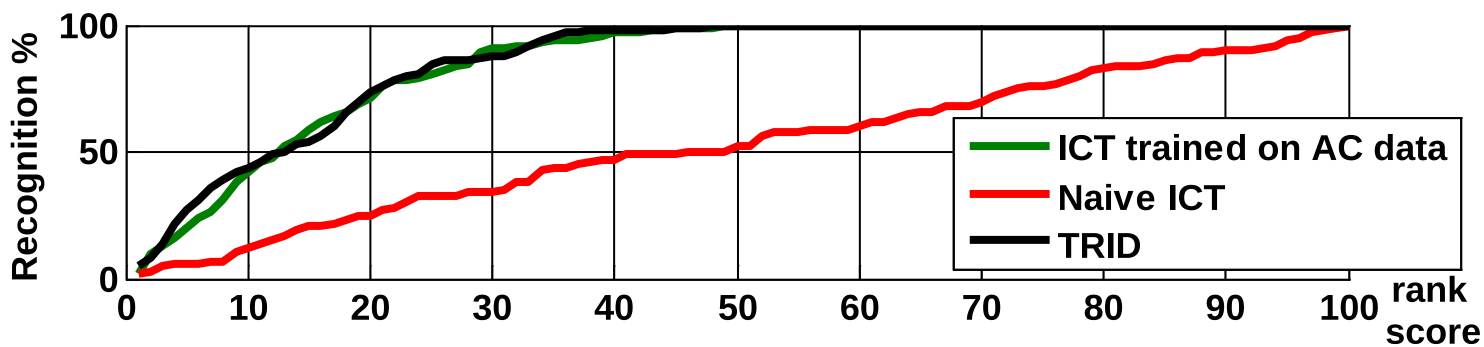


Naive ICT *ICT trained on AC data* *TRID*

Synthetic Experiment 2





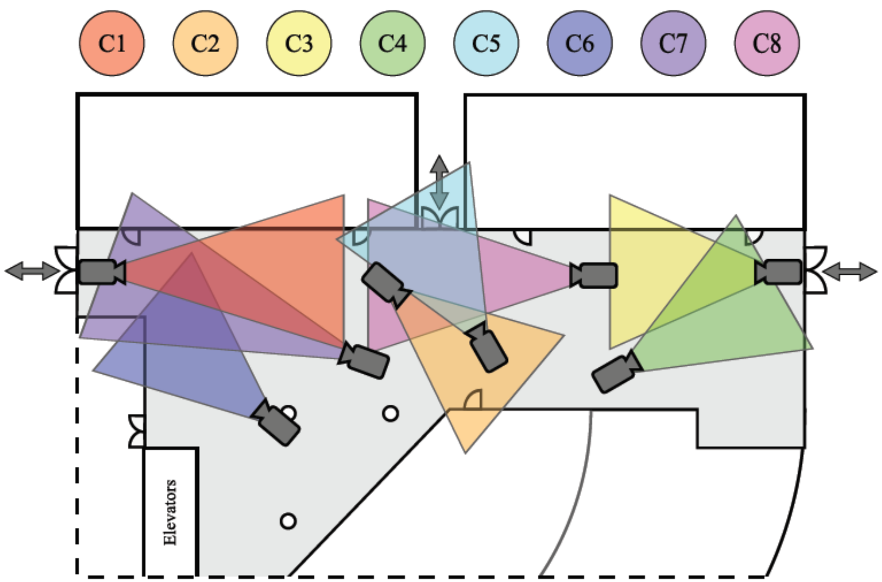


$P(Y_{AC}|x_A, x_C)$:

Naive ICT | *ICT trained on AC data* | *TRID*

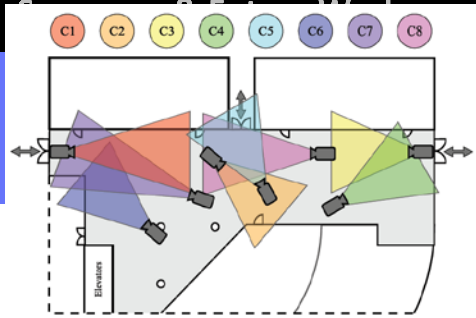
SAIVT-SoftBio Experiment

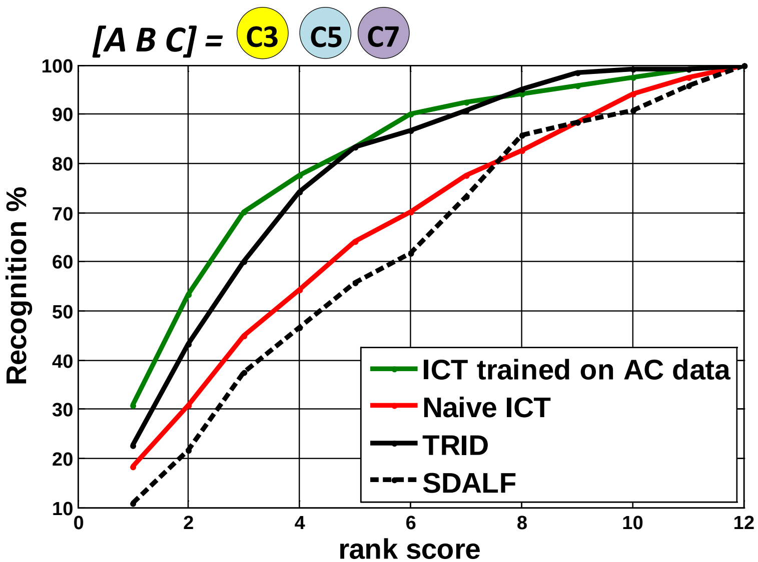
- Testing TRID requires an annotated dataset associated with at least 3 stationary cameras.
- Common ReID benchmark datasets (VIPeR , CAVIAR4REID, iLIDs MCTS, ETHZ) are unsuitable for our set-up.
- Multi-camera surveillance database named SAIVT-SoftBio presented by Bialkowski et al:

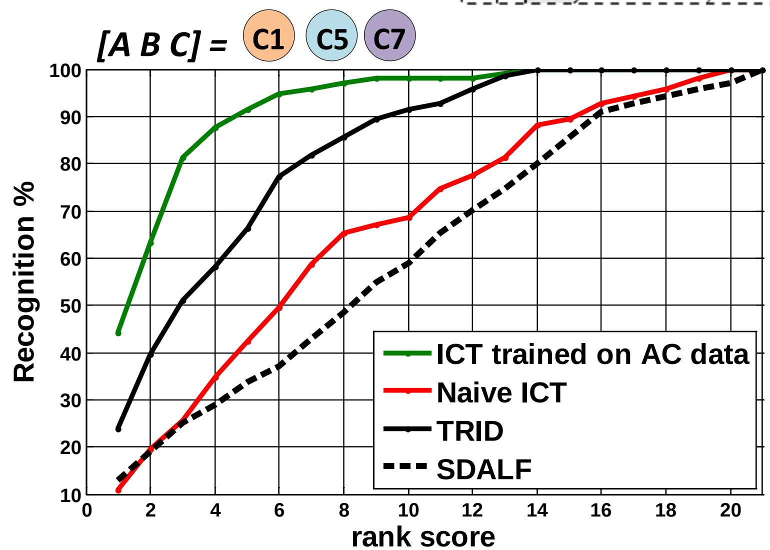


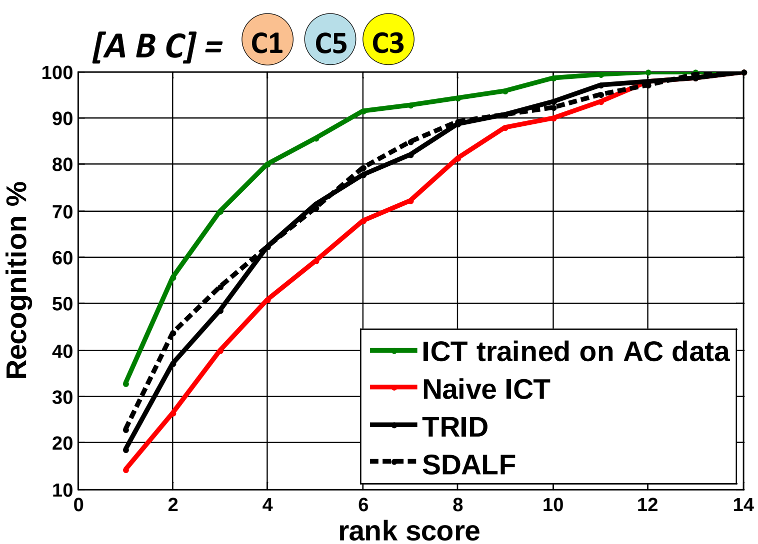
- **image from:** A. Bialkowski, S. Denman, P. Lucey, S. Sridharan, and C. B. Fookes. "A database for person re-identification in multi-camera surveillance networks." (DICTA 2012)

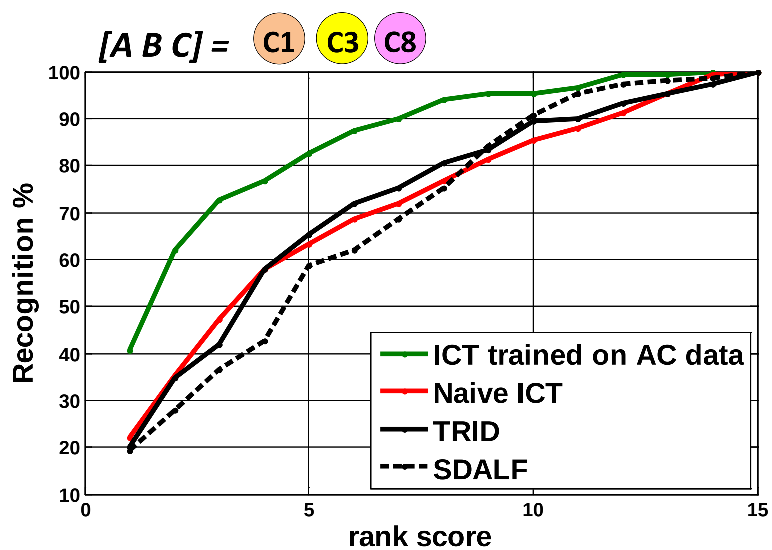
SAIVT-SoftBio Experiment



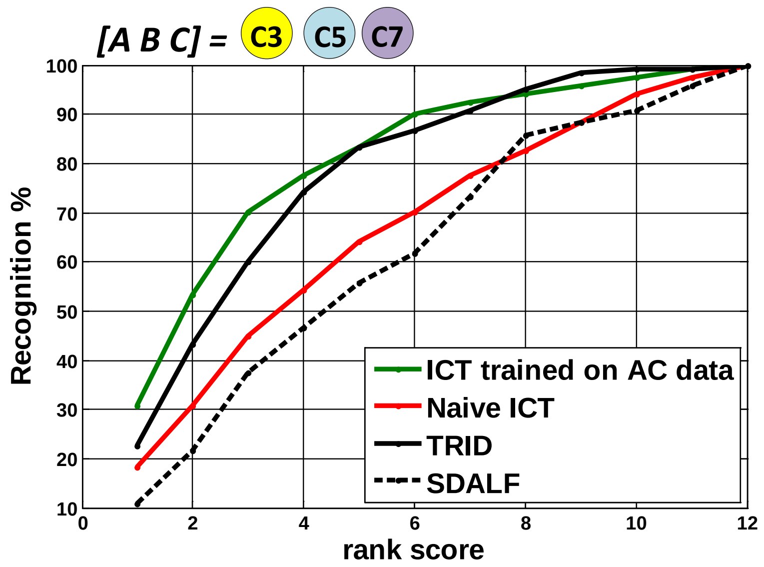


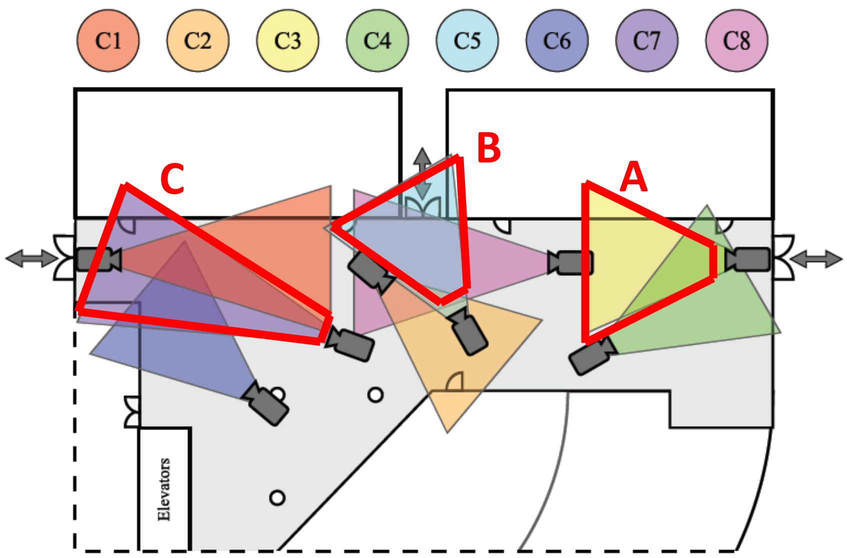




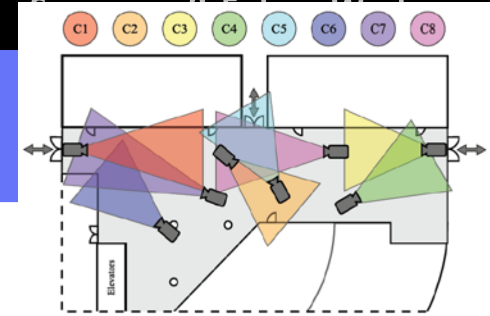


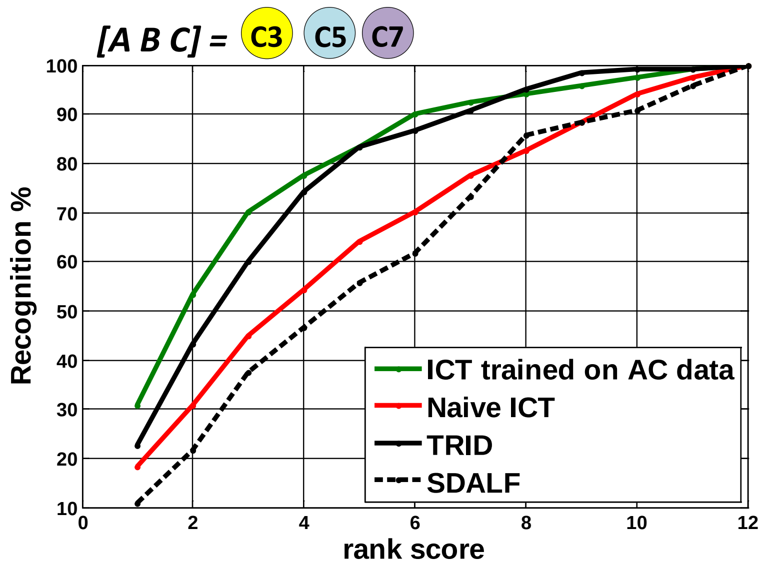
SAIVT-SoftBio Experiment

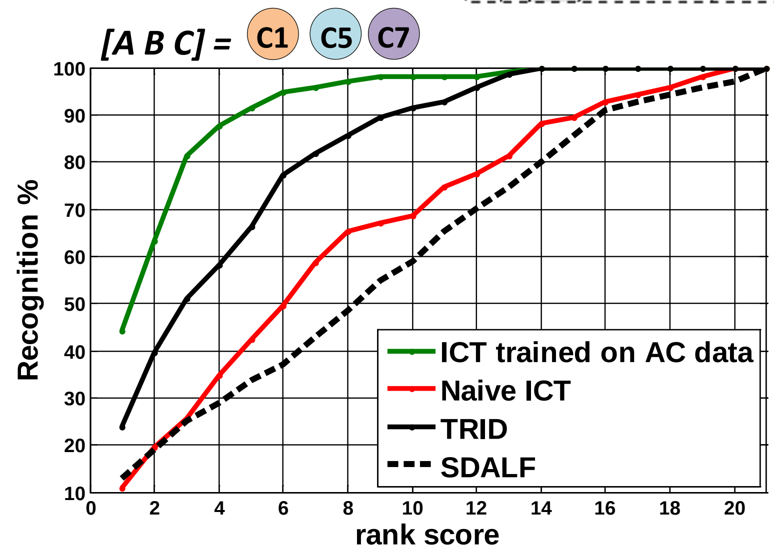


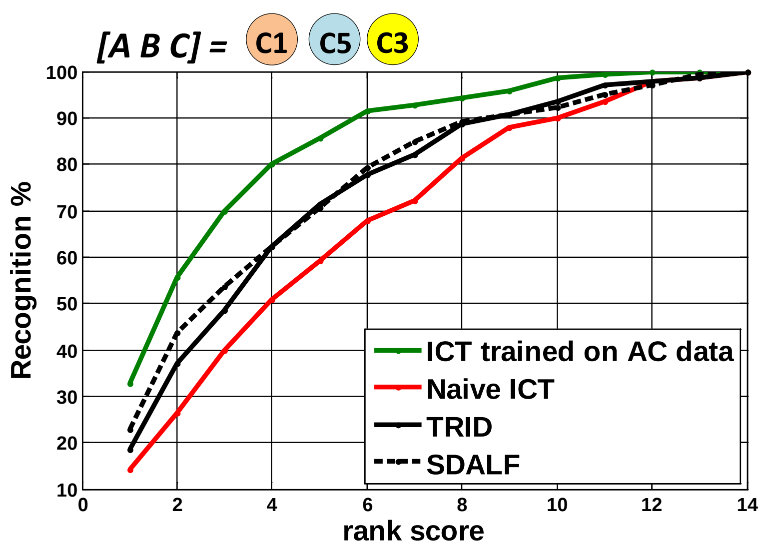


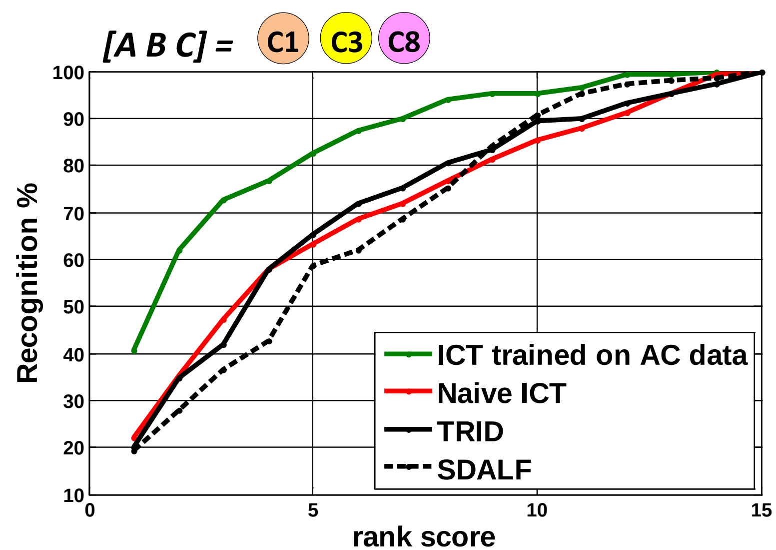
SAIVT-SoftBio Experiment



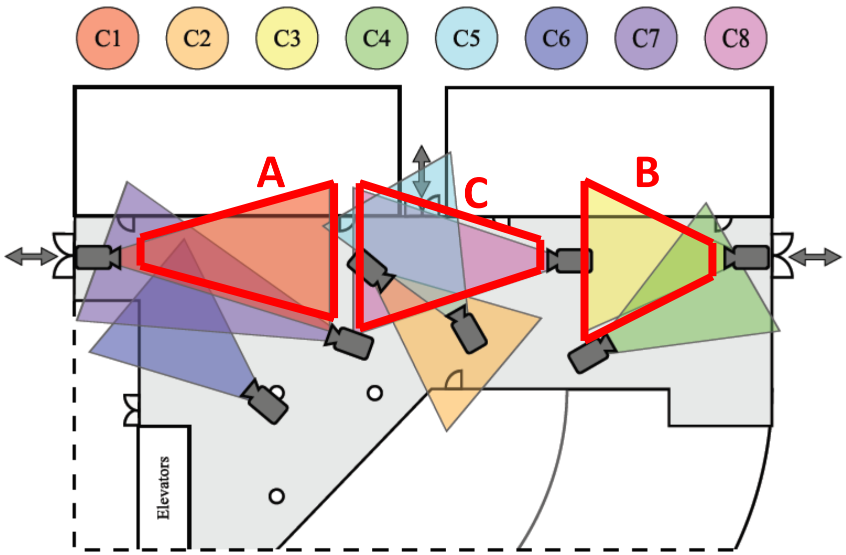


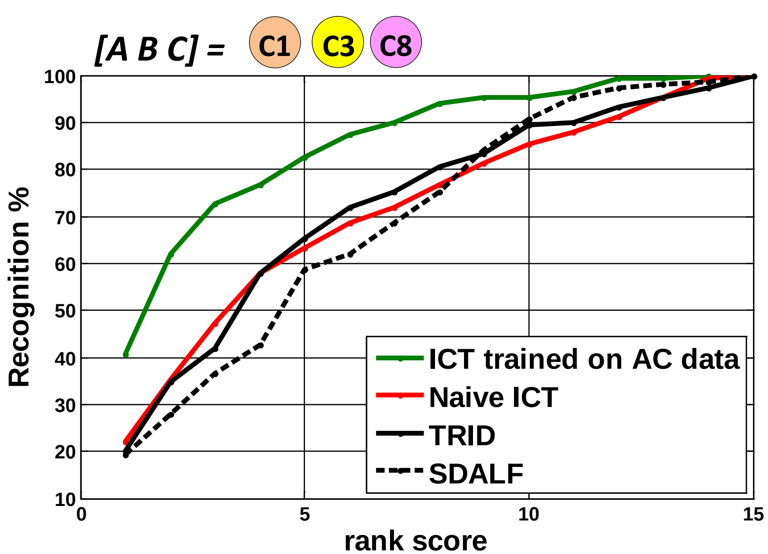






SAIVT-SoftBio Experiment





Summary

- When performing ReID in multi-camera site, we do not have to choose between the less accurate direct method and the computationally expensive learning based method.
  We may perform limited learning, such that requires reasonable resources.
  We have shown that transitively using such learning performs well for the ReID task.

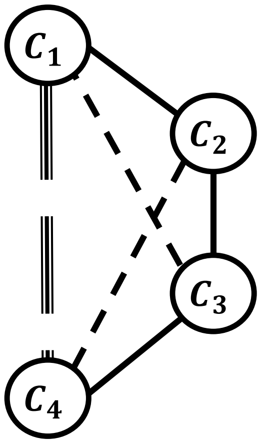
- A specific algorithm based on marginalization was suggested: the TRID - which presents a new approach of transitivity in ReID.

- The accuracy of the TRID is superior to that of a state-of-the-art non learning based approach.

- The TRID algorithm is in fact a general framework that may be combined with different probabilistic classifiers (not necessarily ICT) and of course different features.

Future Work

- Improve the approximation of the transitive integral.
- Transitively use indirect training sets to strengthen direct learning when the set of direct annotated pairs is small, but not empty.
- Test the effect of recursively applying the TRID algorithm, and to study the deterioration of the performance as a function of the number of cameras.
- Generalize the TRID algorithm beyond the scope of ReID to other domain adaptation tasks.



Thank-You.

The Matlab source code of TRID as well as of the ICT are available at:
http://www.cs.technion.ac.il/~tammya/Reidentification.html .