

PERSON RE-IDENTIFICATION

HAVSS 2012

INRIA Sophia Antipolis, the 5th of October

Sławomir BĄK

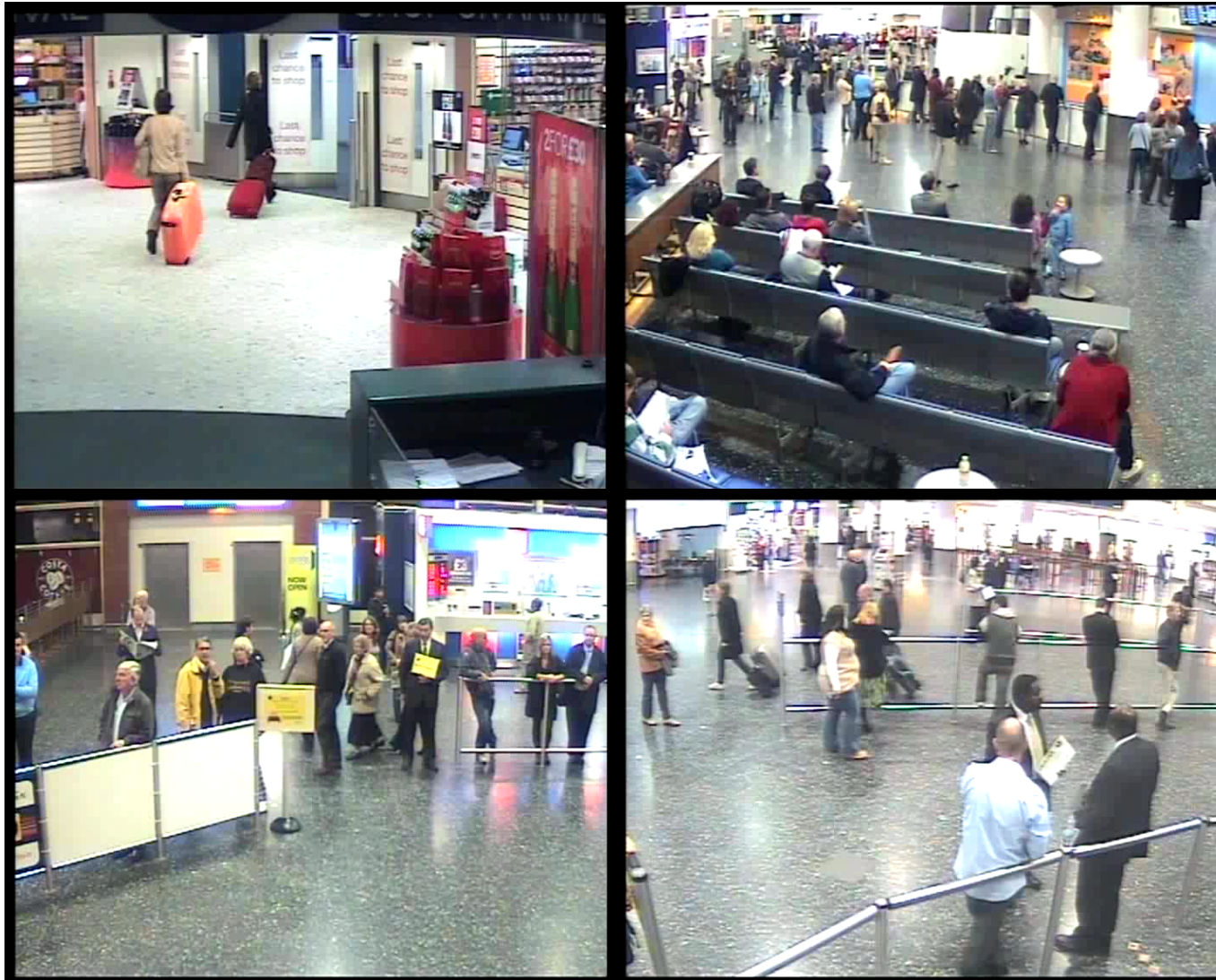
Slawomir.Bak@inria.fr



informatics mathematics
inria

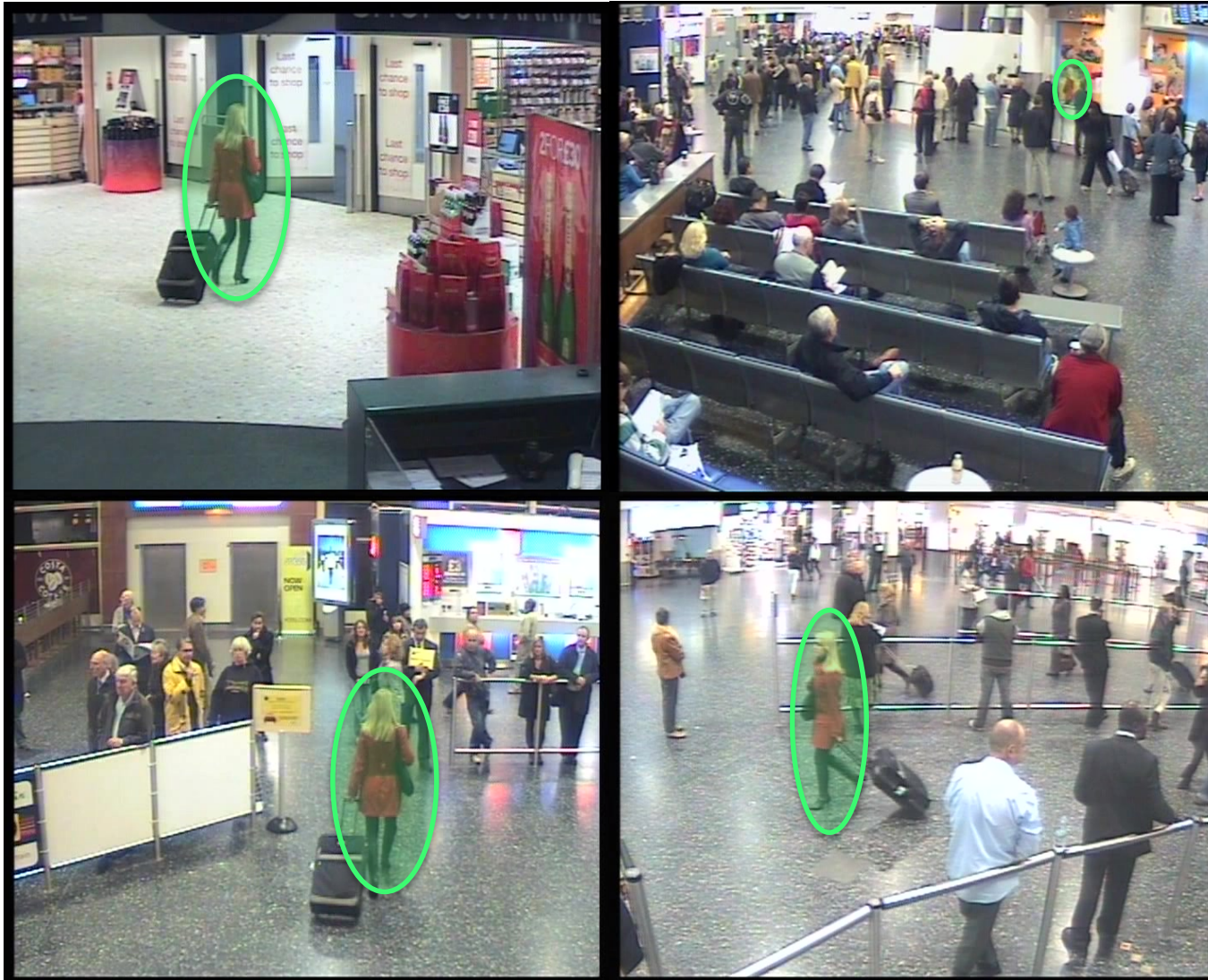


MOTIVATION



CCTV cameras, UK: 4M, London: 1.8M
Human can not perform efficient surveillance after 12minutes

MOTIVATION



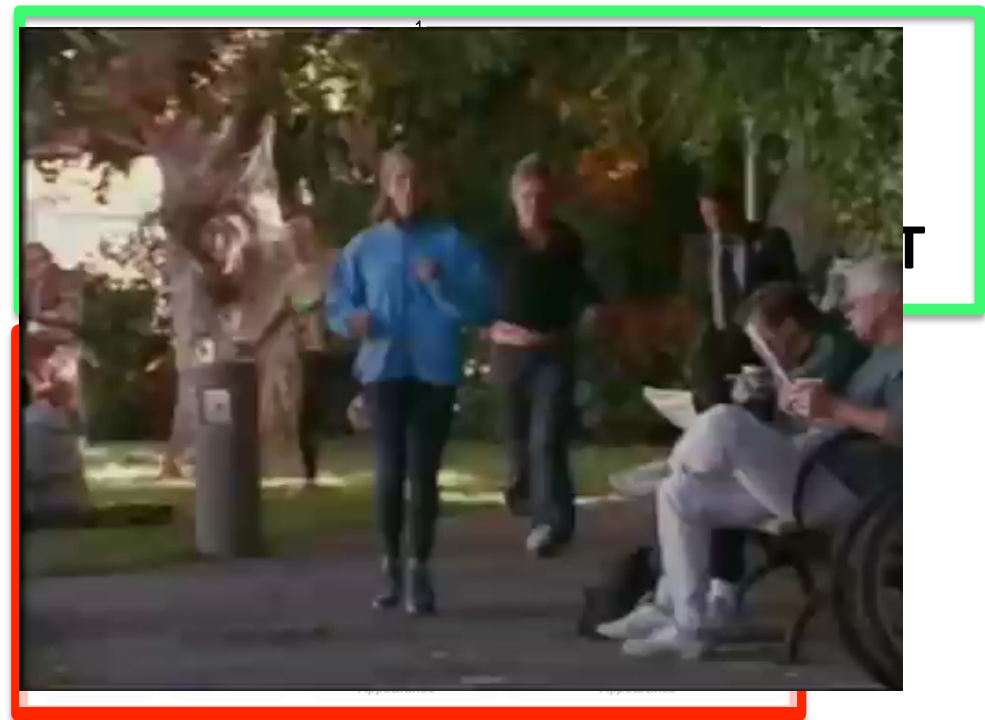
Person **RE-IDENTIFICATION** Problem



THE OBJECTIVE IS TO **DETERMINE** WHETHER A GIVEN **PERSON OF INTEREST** HAS ALREADY BEEN OBSERVED OVER A **NETWORK OF CAMERAS**

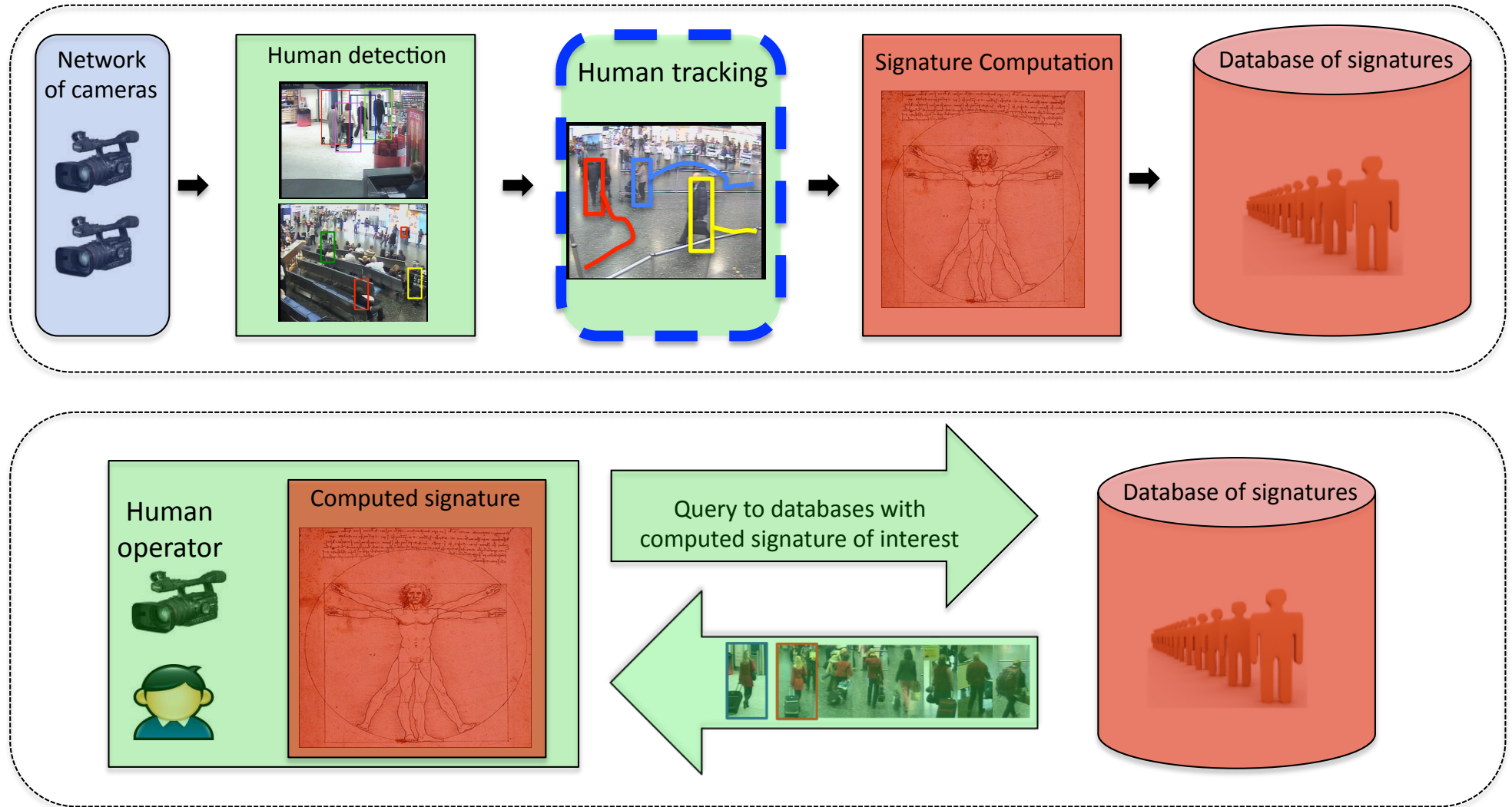


Person **RE-IDENTIFICATION**



Global Appearance

How?

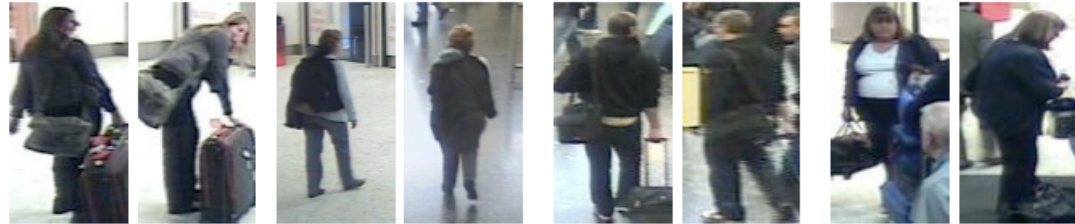


MAIN CHALLENGES

COLOR



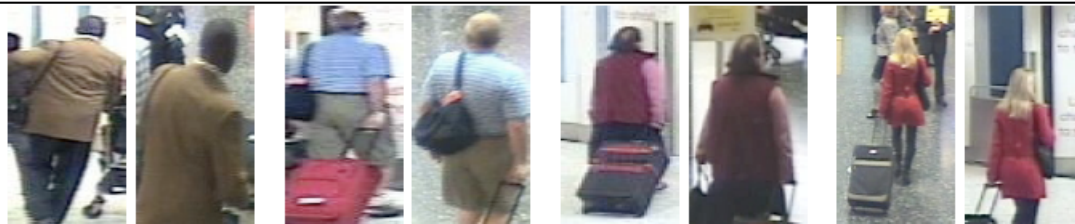
VIEWPOINT



OCCCLUSION



DETECTION



DISCRIMINATIVE FEATURES

Inspired by **human memory** and in particular – **recognition memory**





Outline



Related work



Re-identification using a **single** image

– SCR signature



Re-identification using **multiple** images

– COSMATI signature



Evaluation



Conclusion

RE-IDENTIFICATION



Single-shot



Multiple-shot

Feature-oriented

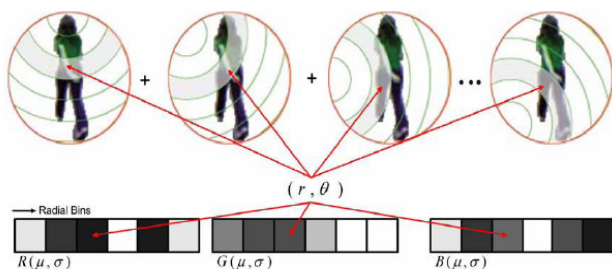
Learning – Discriminative / Metric

Feature-Oriented

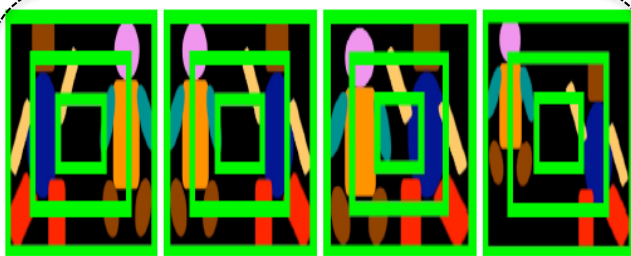
FO



Single-shot



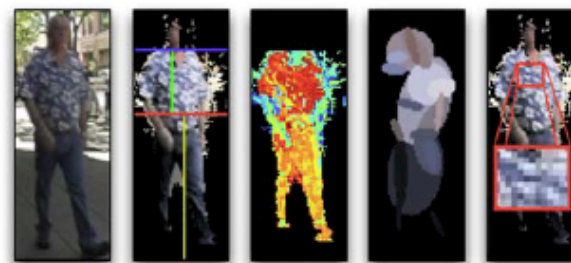
Kang [ICPR 2008]
viewpoint invariant features



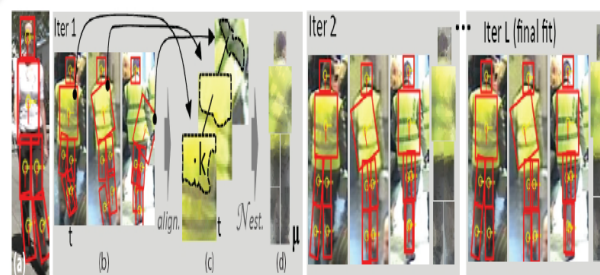
Zheng [BMVC 2009]
Group context – rectangular ring regions



Multiple-shot



Farenzena [CVPR 2010]
Symmetry and asymmetry cues



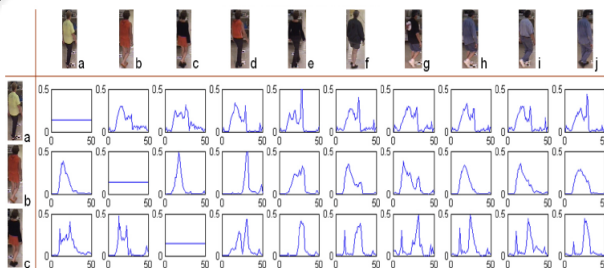
Cheng [BMVC 2011]
Precise body part detections

Learning - Discriminative Methods

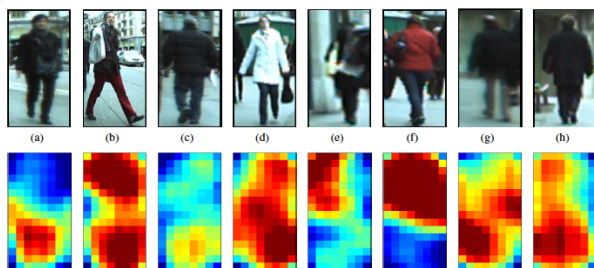
DM



Single-shot



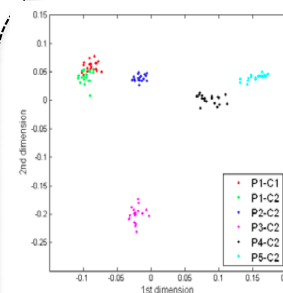
Lin [ISVS 2008]
Dissimilarity Profiles



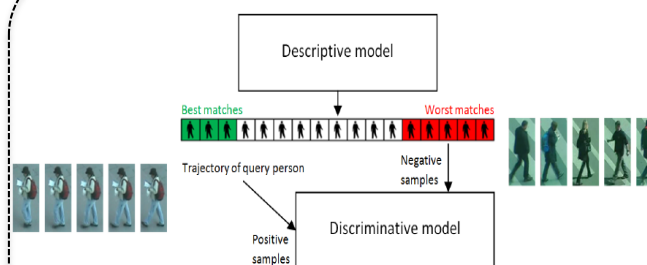
Schwartz [SIBGRAPI 2009]
Discriminative Appearances



Multiple-shot



Truong Cong [ICIAP 2009]
Dimensionality reduction

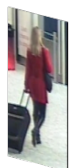


Hirzer [SCIA 2011]
Descriptive + Discriminative model

FO+DM

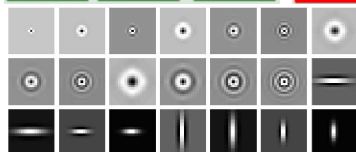
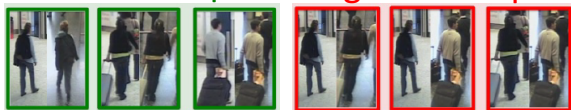
Metric Learning

ML



Single-shot

Positive samples Negative samples



+Boosting

Gray [ECCV 2008]

Ensemble of Localized Features (ELF)

$$P(f(\mathbf{x}_i^p) < f(\mathbf{x}_i^n)) = (1 + \exp \{f(\mathbf{x}_i^p) - f(\mathbf{x}_i^n)\})^{-1}$$

$$f = \arg \min_f r(f, \mathbb{O}),$$

$$r(f, \mathbb{O}) = -\log(\prod_{\mathbb{O}_i} P(f(\mathbf{x}_i^p) < f(\mathbf{x}_i^n))).$$

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{M} \mathbf{x}, \mathbf{M} \succeq 0$$

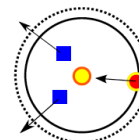
Zheng [CVPR 2011]

Probabilistic model

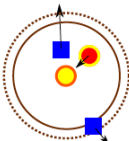
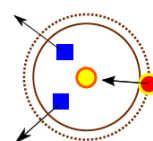


Prosser [BMVC 2010]

Re-identification by ranking



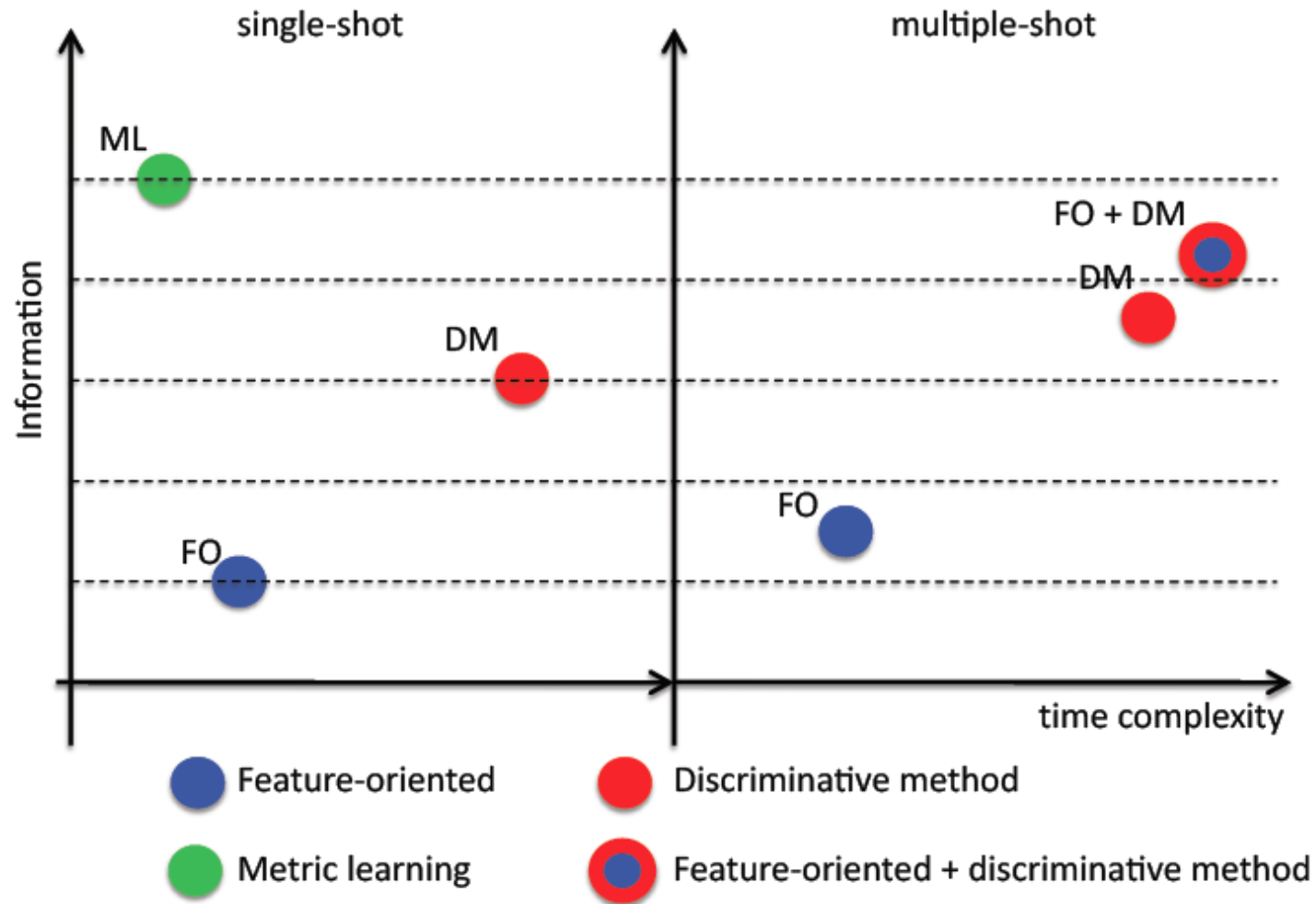
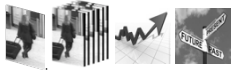
LMNN

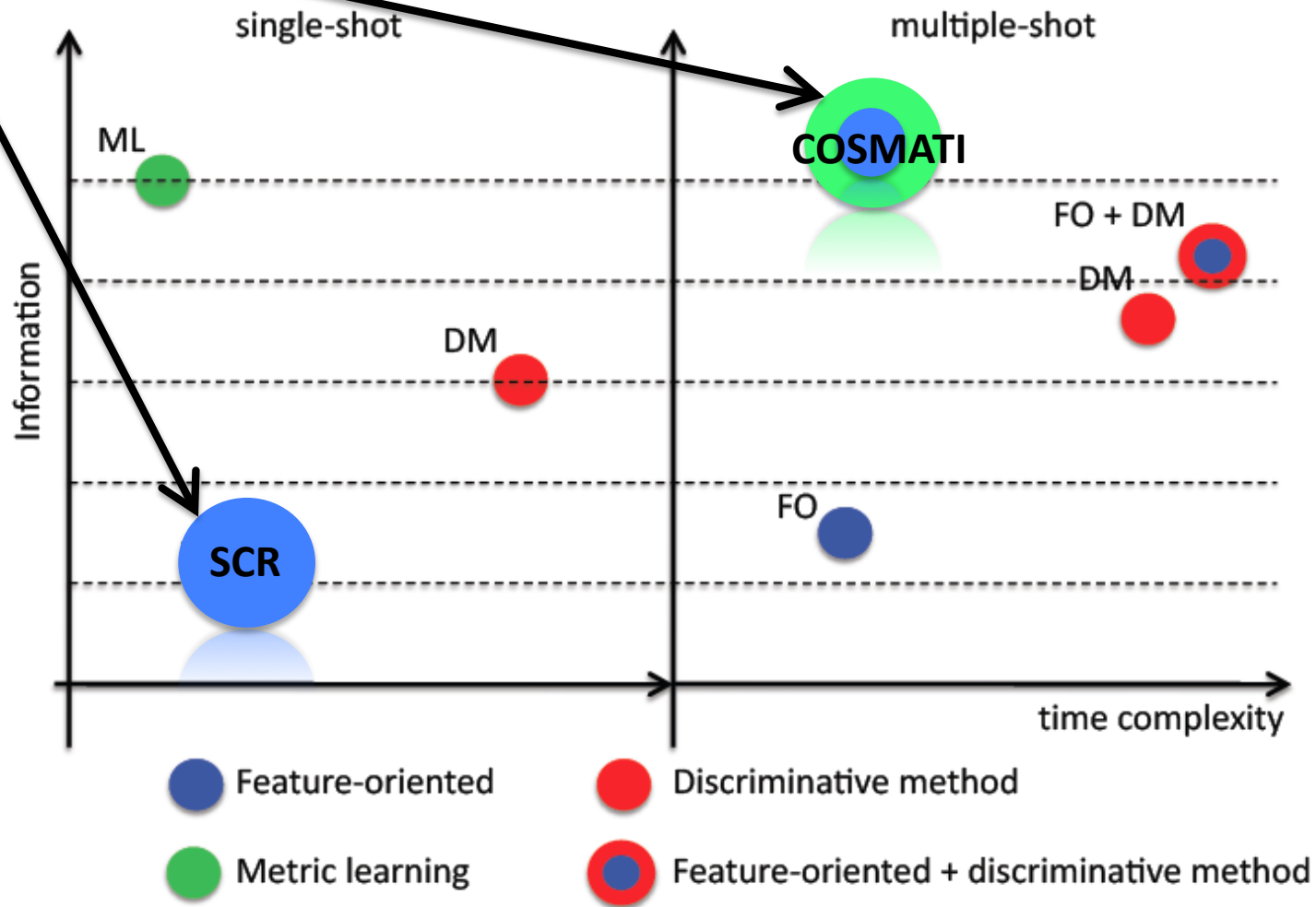


LMNN-R

Dikmen [ACCV 2010]

Large Margin Nearest Neighbor







SINGLE-SHOT APPROACH

Spatial Covariance Regions (SCR)

Image



Body Part
Detections

Body Parts



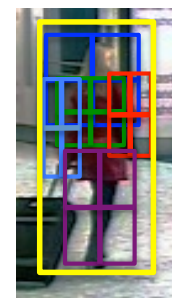
Color
Normalization

Normalized Image



Signature
Computation

SCR Signature

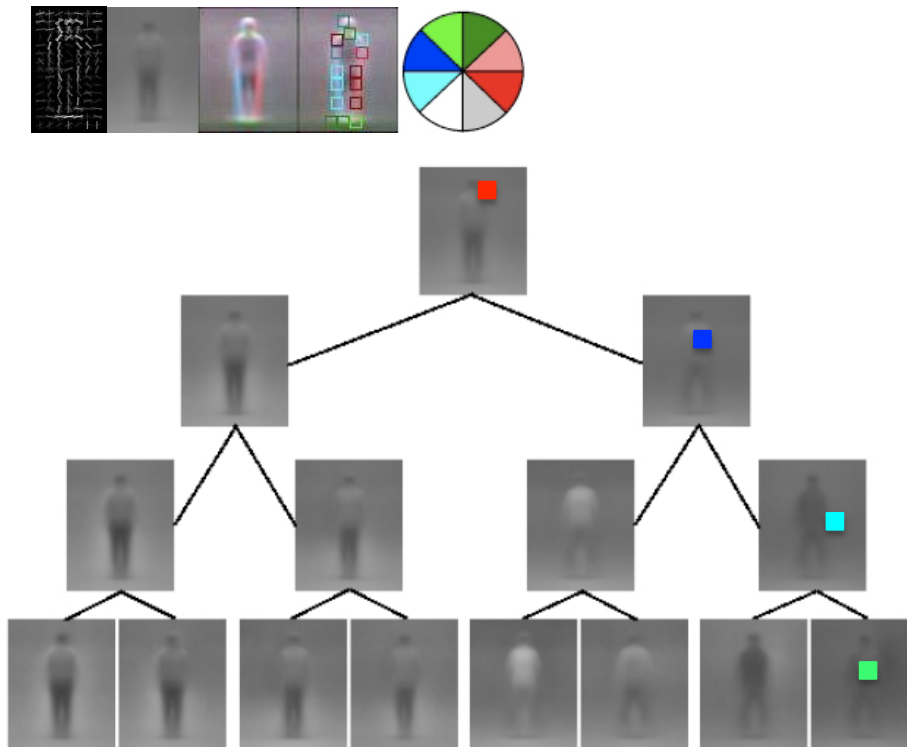




SINGLE-SHOT APPROACH

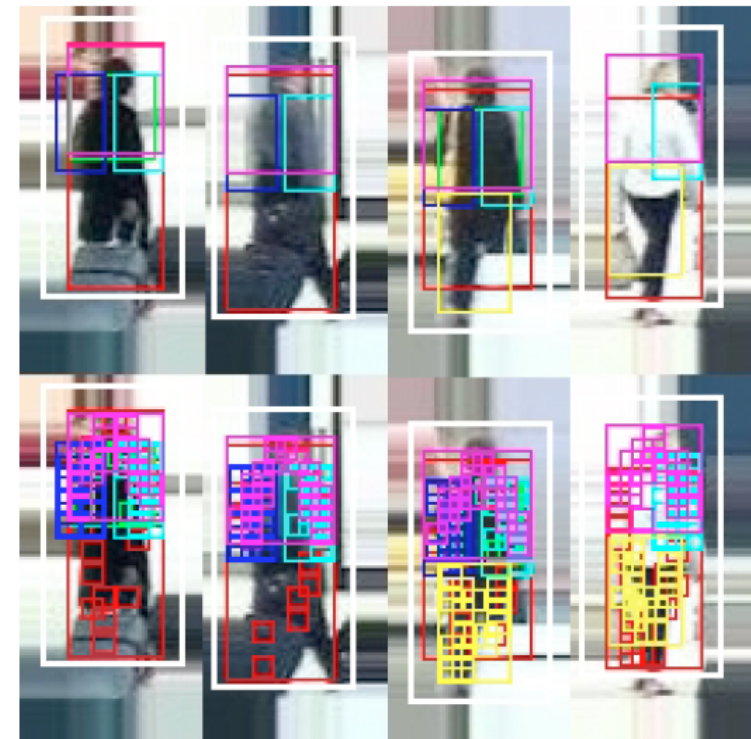
Body part detections

Tree classifier, **Corvee** [AVSS 2010]



The **tree** of **HOG** cells

Each **body part** has own tree classifier based on **HOG**



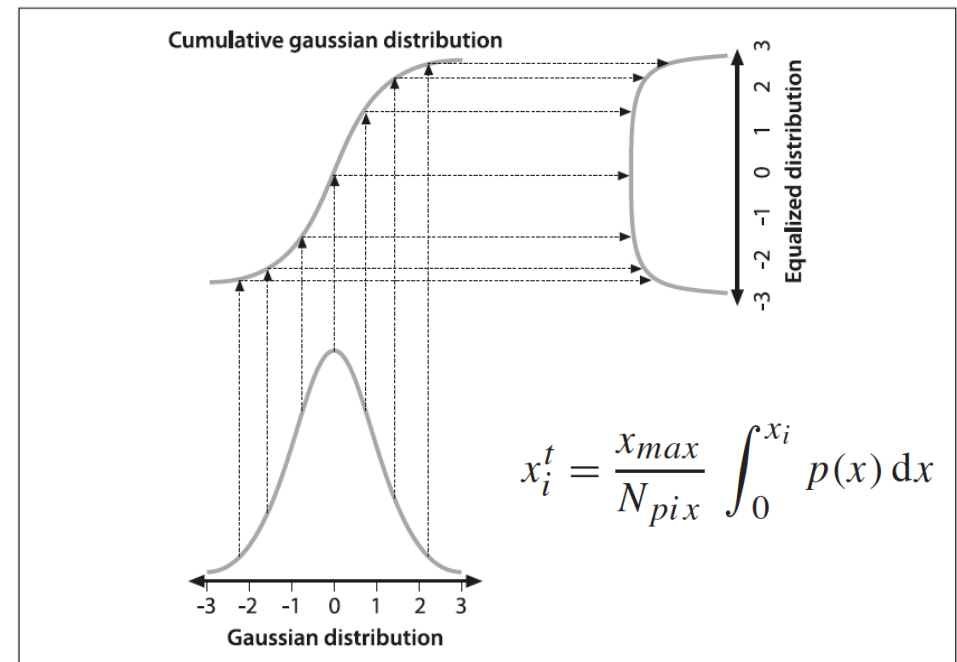
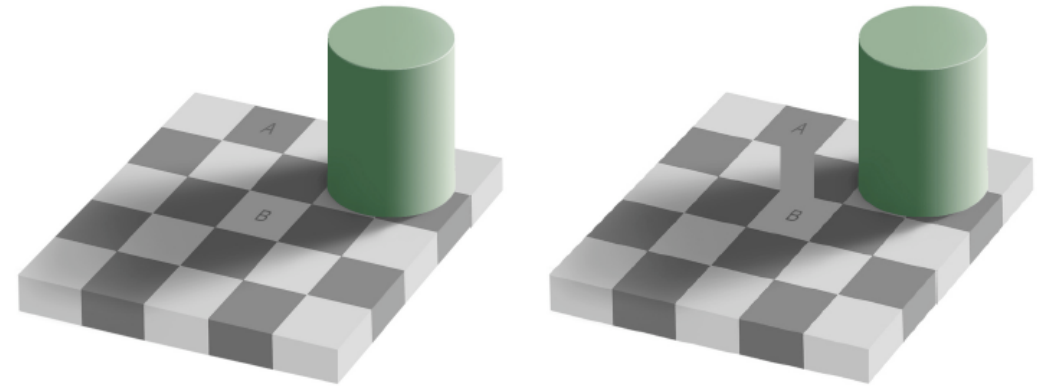
Examples of detected body parts
and the corresponding **HOG** cells



SINGLE-SHOT APPROACH

Color normalization

- *Histogram Equalization*



IN EACH COLOR CHANNEL (RGB)

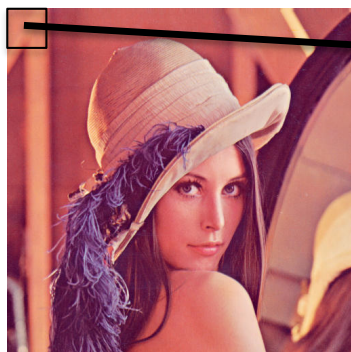


SINGLE-SHOT APPROACH

Signature computation

Covariance descriptor

O. Tuzel, F. Porikli, and P. Meer. *Region covariance: A fast descriptor for detection and classification* [ECCV 2006]



$$\mathbf{f}_k = [f_{k,1}, \dots, f_{k,d}]$$

$$C_R = \frac{1}{n-1} \sum_{k=1}^n (\mathbf{f}_k - \mu)(\mathbf{f}_k - \mu)^T$$

$$C_R(i, j) = \frac{1}{n-1} \sum_{k=1}^n (f_k(i) - \mu(i))(f_k(j) - \mu(j))$$

$$\begin{pmatrix} E[(f(1) - \mu(1))(f(1) - \mu(1))] & & & \\ E[(f(2) - \mu(2))(f(1) - \mu(1))] & E[(f(2) - \mu(2))(f(2) - \mu(2))] & & \\ \vdots & & \ddots & \\ E[(f(d) - \mu(d))(f(1) - \mu(1))] & & & E[(f(d) - \mu(d))(f(d) - \mu(d))] \end{pmatrix}$$

$d = 11$

LOCALIZATION COLOR

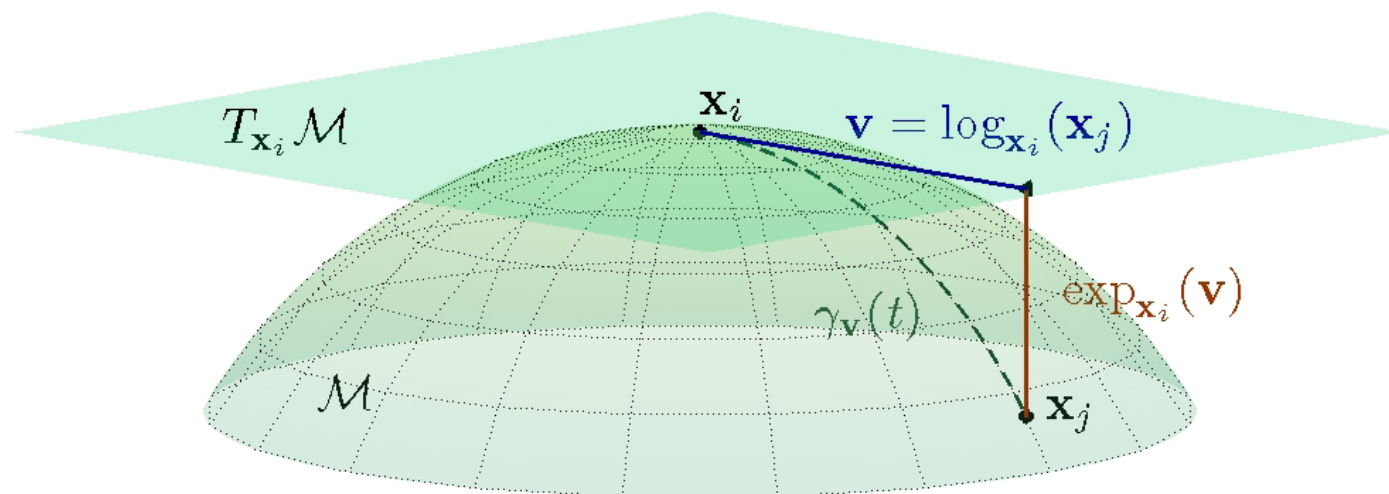
GRADIENT

$$\mathbf{f}_k = \left[\underbrace{x, y}_{\text{LOCALIZATION}}, \underbrace{R_{xy}, G_{xy}, B_{xy}}_{\text{COLOR}}, \underbrace{\|\nabla_{xy}^R\|, \theta_{xy}^R, \|\nabla_{xy}^G\|, \theta_{xy}^G, \|\nabla_{xy}^B\|, \theta_{xy}^B}_{\text{GRADIENT}} \right]$$



Signature computation

Covariance descriptor



Riemannian manifold

$$\rho(C_i, C_j) = \sqrt{\sum_{k=1}^d \ln^2 \lambda_k(C_i, C_j)}$$

the generalized eigenvalue problem ($C_i x = \lambda C_j x$)

W. Förstner and B. Moonen. A metric for covariance matrices.

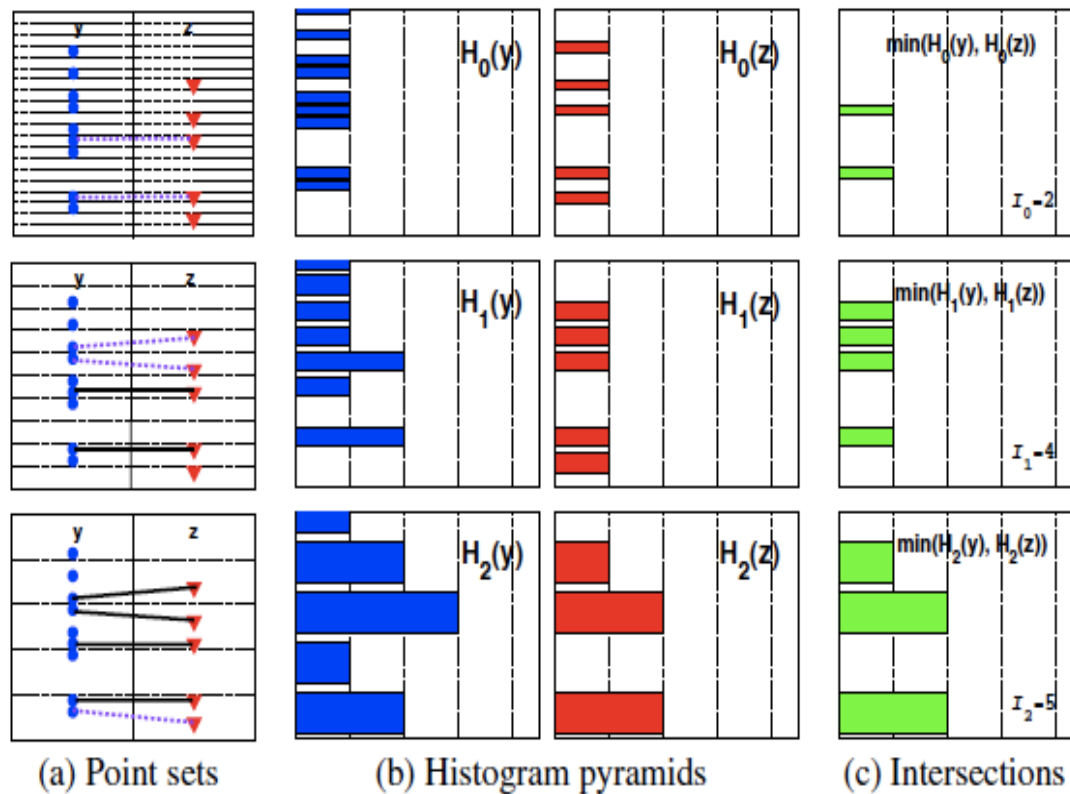


SINGLE-SHOT APPROACH

Spatial Pyramid Matching

K. Grauman and T. Darrell, ICCV 2005

The pyramid match kernel: Discriminative classification with sets of image features



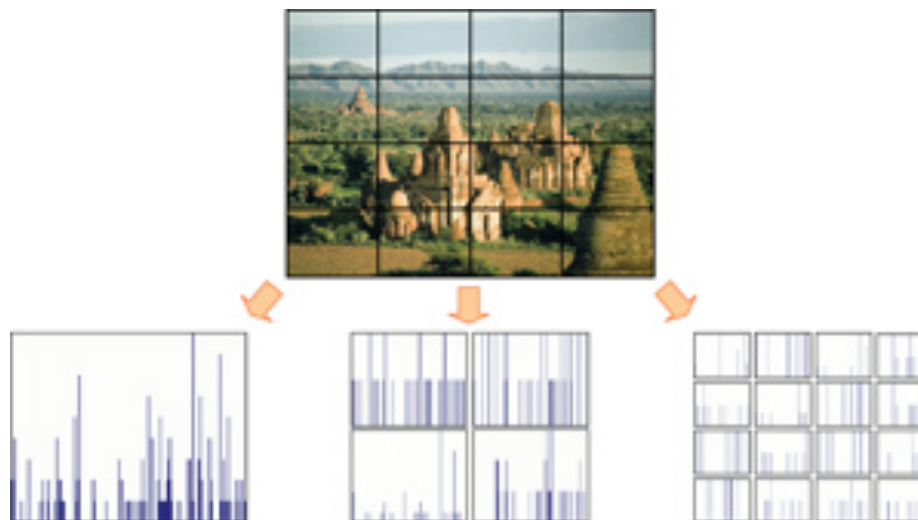


SINGLE-SHOT APPROACH

Spatial Pyramid Matching

S. Lazebnik, C. Schmid, and J. Ponce, CVPR 2006

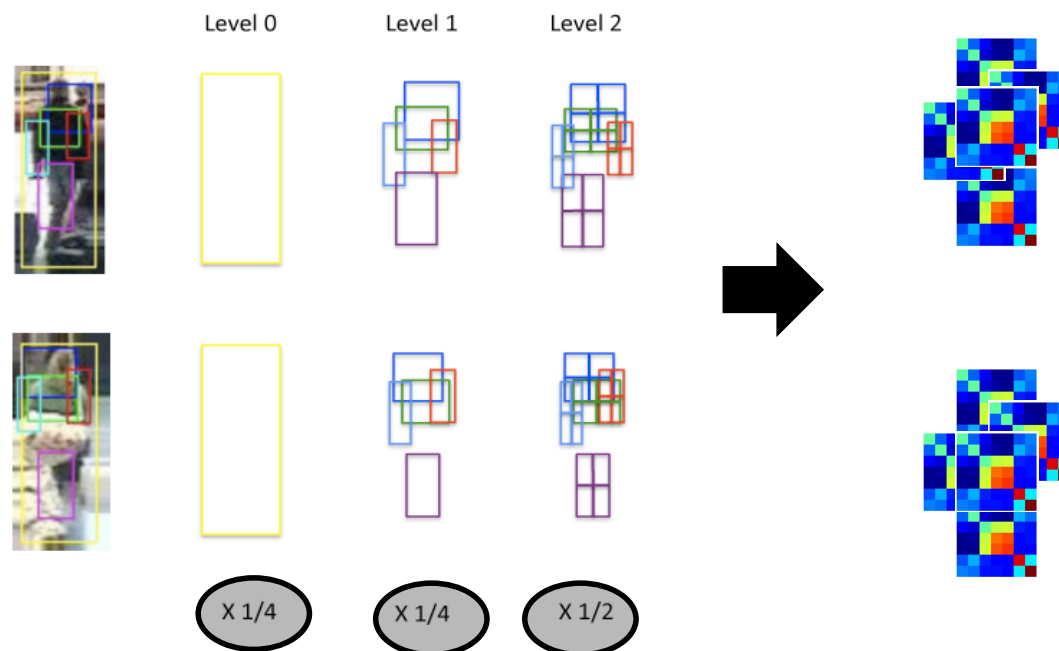
Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories





SINGLE-SHOT APPROACH

Spatial Pyramid Matching



$$\mathcal{D}(I_i, I_j) = \kappa^L(I_i, I_j).$$

$$\kappa^L(X, Y) = \mathcal{I}^L + \sum_{l=0}^{L-1} \frac{1}{2^{L-l}} (\mathcal{I}^l - \mathcal{I}^{l+1}) = \frac{1}{2^L} \mathcal{I}^0 + \sum_{l=1}^L \frac{1}{2^{L-l+1}} \mathcal{I}^l,$$

$$\mathcal{D}^l = \frac{\sum_{k=1}^{n(l)} \rho(C_k^i, C_k^j) - \sum_{m \in \mathcal{M}_z} \rho_m}{n(l) - z}$$

largest dissimilarities:
robust to

occlusions

$$\mathcal{I}^0 = \frac{1}{\rho(C_i, C_j)}, \quad \mathcal{I}^l = \frac{1}{\mathcal{D}^l}$$

geodesic distance

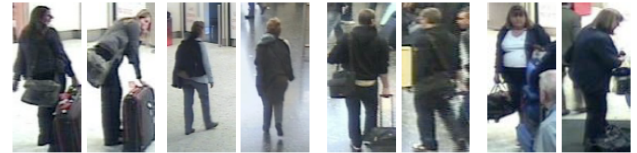


SINGLE-SHOT APPROACH

Spatial Covariance Regions (SCR)

COLOR

Histogram equalization + Covariance features

VIEWPOINT

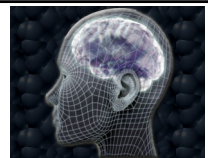
Detection of body parts:
addresses *the correspondence problem*

OCCCLUSION

Dissimilarity measure

DETECTION

Detection of body parts

DISCRIMINATIVE FEATURES

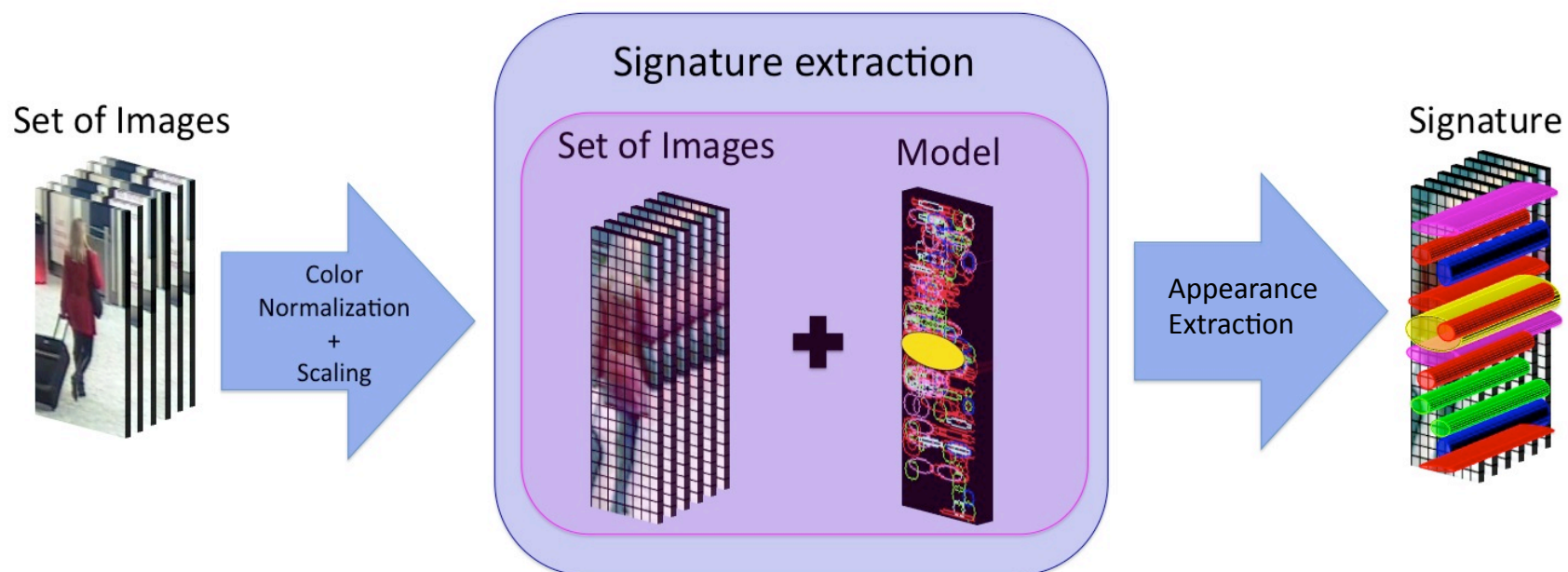
Spatial pyramid matching?

- (1) Dependent on accuracy of body part detections
- (2) Computationally heavy distance operator (11 x 11 covariance)
- (3) No discriminative features



MULTIPLE-SHOT APPROACH

COrrelation-based ***S***election of covariance ***MAT***rices (***COSMATI***)

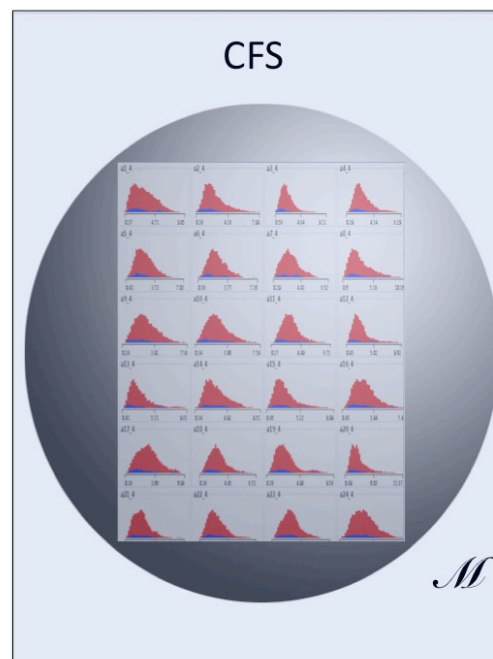




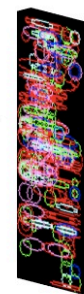
MULTIPLE-SHOT APPROACH

*CO*rrelation-based *S*election of covariance *MAT*rices (*COSMATI*)

Learning a model



Model





MULTIPLE-SHOT APPROACH

Feature space

Covariance descriptor

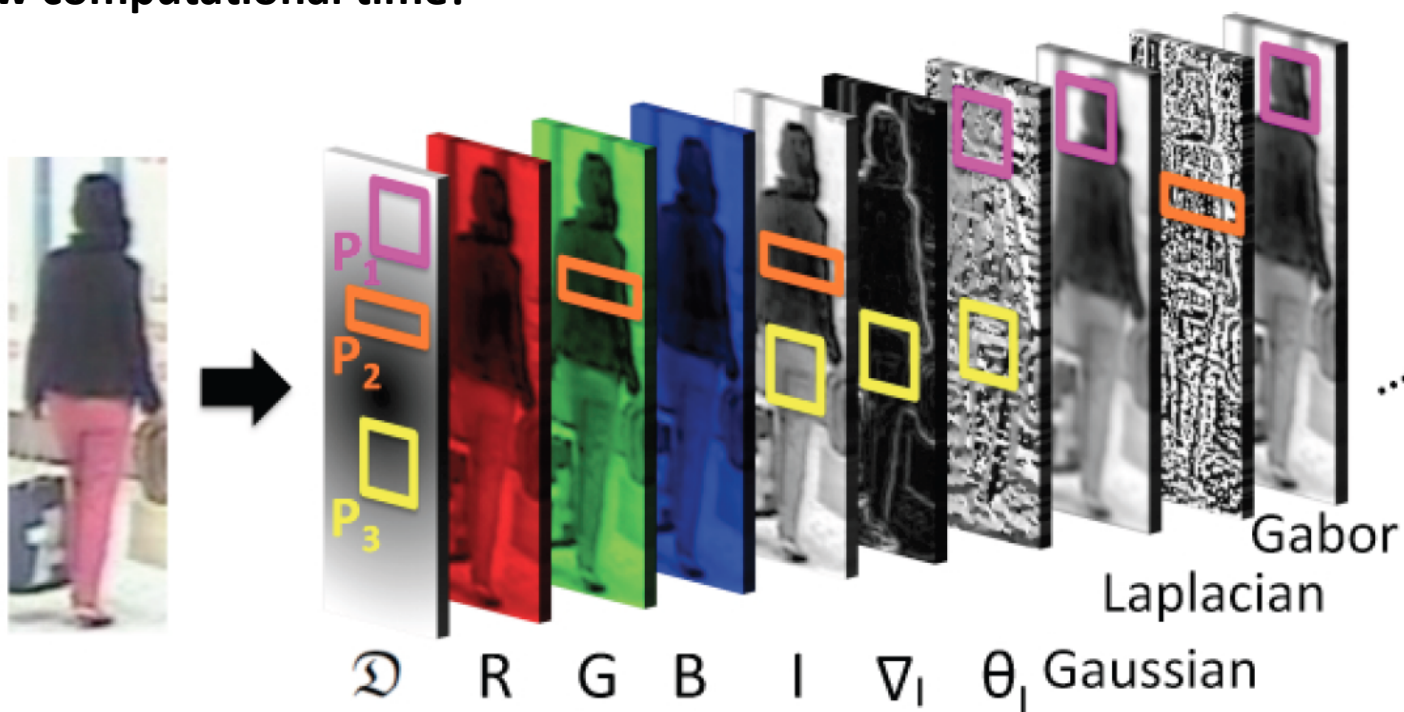
$f_k = [f_{k,1}, \dots, f_{k,d}]$, **NO** *a priori* feature

The idea is to characterize **different regions**
of an object by extracting **different kinds of features**

Feature layers: color, gradients and filter responses: $L = \{R, G, B, I, \nabla_I, \theta_I, \dots\}$

We select **4x4** covariance matrices: $\mathcal{C} = \{(P, \text{cov}(\mathcal{D}, l_i, l_j, l_k)) : l_i, l_j, l_k \in L; P \in \mathcal{P}\}$

Low computational time!





MULTIPLE-SHOT APPROACH

Learning by correlations in a covariance metric space

 a^c $a^{c'}$

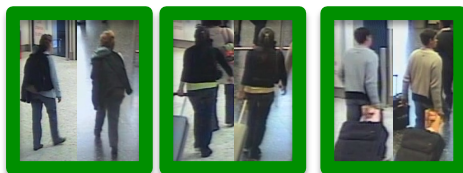
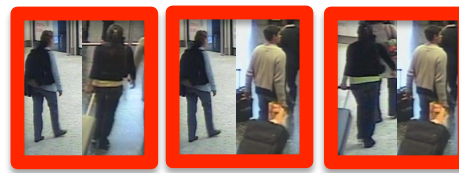
$$\mathcal{C} = \{ (P, cov(\mathcal{D}, l_i, l_j, l_k)) : l_i, l_j, l_k \in L; P \in \mathbf{P} \}$$

Size of $\mathcal{C}$ is n

$$\delta(a_{i,j}^c, a_{k,l}^{c'}) = [\rho(a_{i,j}^c[z], a_{k,l}^{c'}[z])]_{z=1 \dots n}^T$$

"A good feature subset is one that contains features **highly correlated** with (predictive of) **the class**, yet **uncorrelated** with (not predictive of) **each other**"

Hall 1999

 δ^+  δ^- 

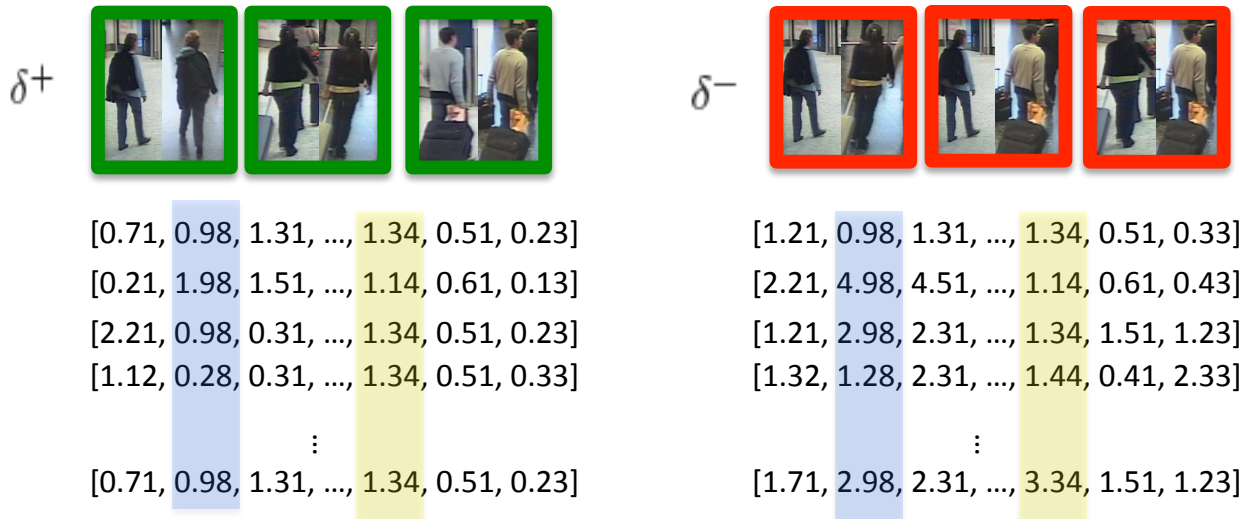
[0.71, 0.98, 1.31, ..., 1.34, 0.51, 0.23]
 [0.21, 1.98, 1.51, ..., 1.14, 0.61, 0.13]
 [2.21, 0.98, 0.31, ..., 1.34, 0.51, 0.23]
 [1.12, 0.28, 0.31, ..., 1.34, 0.51, 0.33]
 ⋮
 [0.71, 0.98, 1.31, ..., 1.34, 0.51, 0.23]

[1.21, 0.98, 1.31, ..., 1.34, 0.51, 0.33]
 [2.21, 4.98, 4.51, ..., 1.14, 0.61, 0.43]
 [1.21, 2.98, 2.31, ..., 1.34, 1.51, 1.23]
 [1.32, 1.28, 2.31, ..., 1.44, 0.41, 2.33]
 ⋮
 [1.71, 2.98, 2.31, ..., 3.34, 1.51, 1.23]



MULTIPLE-SHOT APPROACH

Learning by correlations in a covariance metric space



Correlation-based Feature Selection (CFS) [Hall 1999]

$f_z \in \mathcal{C}$, let X be a nominal valued feature obtained by discretization of f_z [Fayyad IJCAI 1993]

$$H(X) = - \sum_{x \in X} p(x) \log_2 p(x)$$

$$H(X | Y) = - \sum_{y \in Y} p(y) \sum_{x \in X} p(x | y) \log_2 p(x | y)$$

$$Gain = H(X) - H(X | Y)$$

$$= H(Y) - H(Y | X)$$

$$= H(X) + H(Y) - H(X, Y)$$

the information gain (mutual information)



MULTIPLE-SHOT APPROACH

Correlation-based Feature Selection (CFS)

Calculate correlations

feature-class:

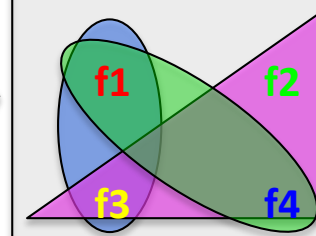
f1 f2 f3 f4

class

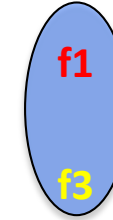
feature-feature:

f1 ↔ f2
f3 ↔ f4
f1 ↔ f3
f1 ↔ f4
f2 ↔ f3
f2 ↔ f4

Best First Search



Feature subset



$$r_{XY} = 2 \times \left[\frac{\text{Gain}}{H(X) + H(Y)} \right] \quad \overline{r_{cf}} = \frac{1}{k} \sum_{f_z \in \mathfrak{S}} r_{cf_z}, \quad \overline{r_{ff}} = \frac{2}{k(k-1)} \sum_{\substack{f_i, f_j \in \mathfrak{S} \\ i < j}} r_{f_i f_j},$$

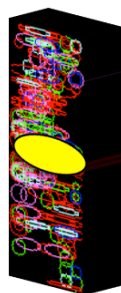
$$\mathfrak{M}(\mathfrak{S}) = \frac{k \overline{r_{cf}}}{\sqrt{k + k(k+1) \overline{r_{ff}}}},$$



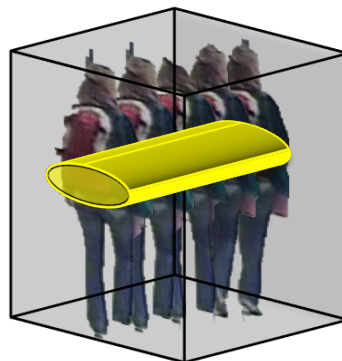
MULTIPLE-SHOT APPROACH

Signature extraction

Model



Set of frames

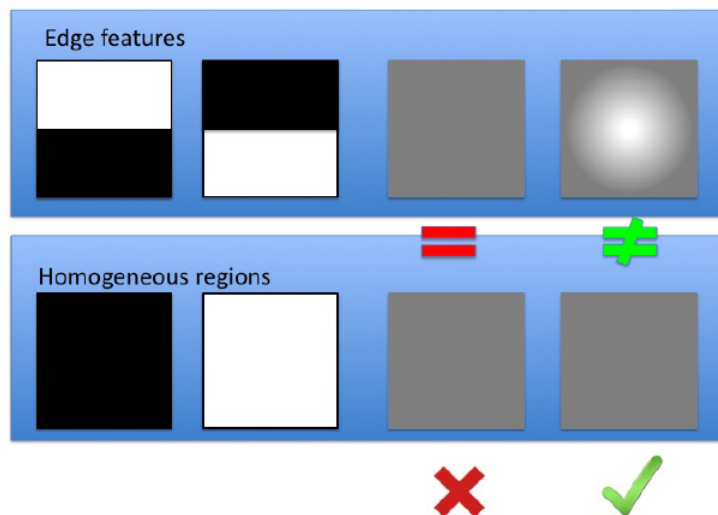
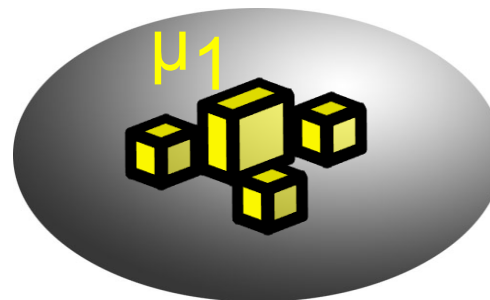


Signature



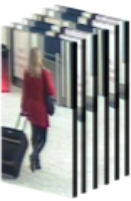
Simple Average

MRC

*Riemannian Manifold***Mean Riemannian Covariance (MRC)**

$$\mu = \arg \min_{C \in \mathcal{M}} \sum_{i=1}^N \rho^2(C, C_i)$$

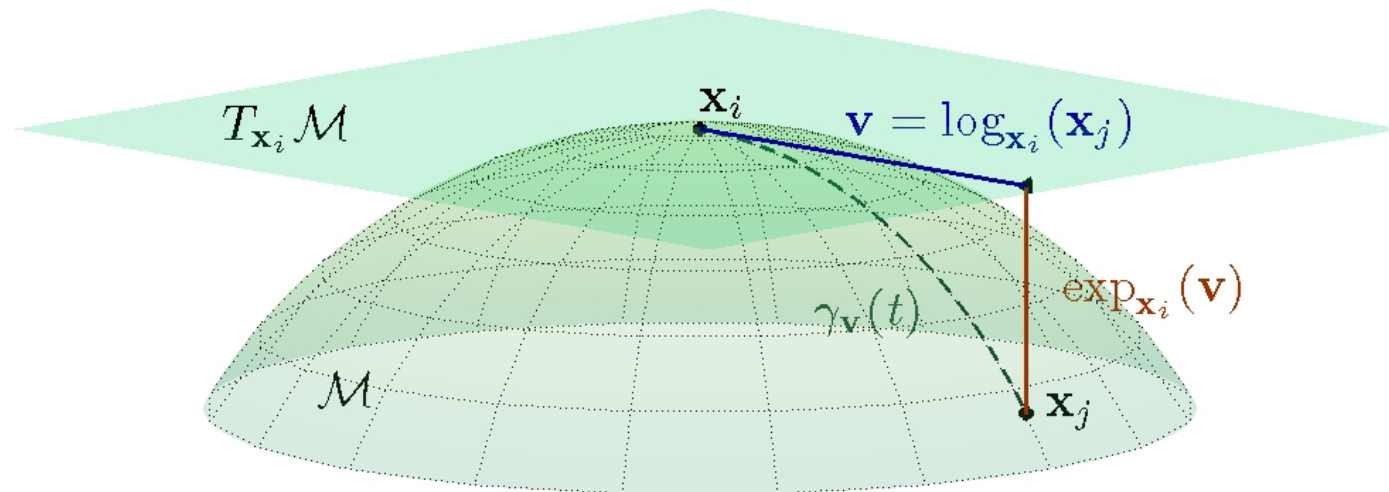
Computing **MRC**, we **blend** appearance information from multiple images, providing partially **pose invariant** representation



MULTIPLE-SHOT APPROACH

Signature extraction

Mean Riemannian Covariance (MRC) computation



An example of a two-dimensional manifold

MRC

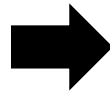
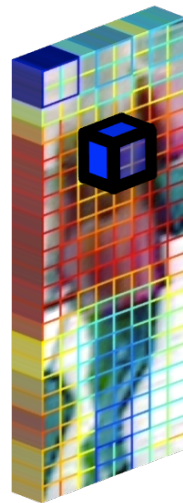
$$\mu = \arg \min_{C \in \mathcal{M}} \sum_{i=1}^N \rho^2(C, C_i) \quad \longrightarrow \quad \mu_{t+1} = \exp_{\mu_t} \left[\frac{1}{N} \sum_{i=1}^N \log_{\mu_t}(C_i) \right]$$

Pennec 2006, A Riemannian Framework for Tensor Computing

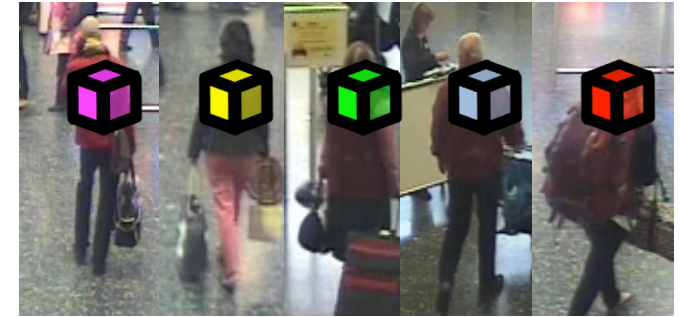
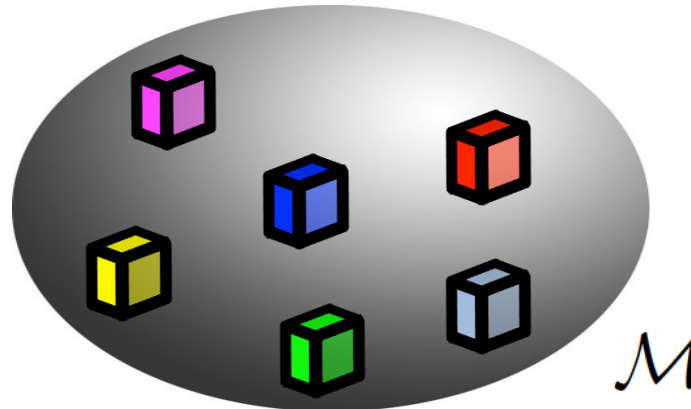


MULTIPLE-SHOT APPROACH

MRC discriminants & distance computation



VARIANCE



VS

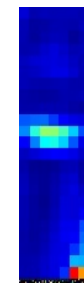
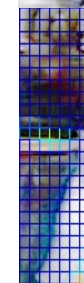
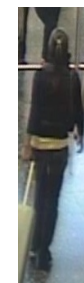
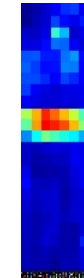
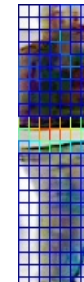


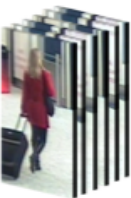
$$\sigma_{i,j}^c = \frac{1}{n-1} \sum_{k=1; k \neq i}^n \rho^2(\mu_{i,j}^c, \mu_{k,j}^c).$$

$\sigma_{i,j}^c \downarrow \longrightarrow$ background / common patterns

$\sigma_{i,j}^c \uparrow \longrightarrow$ the most discriminative patterns

$$S(s_A, s_B) = \sum_{i \in K} \frac{\sigma_{A,i} + \sigma_{B,i}}{\rho(\mu_{A,i}, \mu_{B,i})} / \|K\|$$





MULTIPLE-SHOT APPROACH

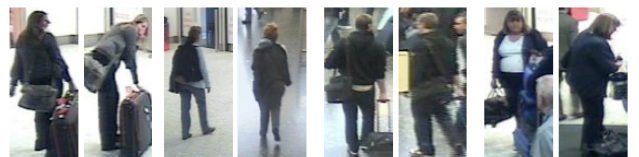
*CO*rrelation-based *S*election of covariance *MAT*rices (*COSMATI*)

COLOR



Histogram equalization + Covariance features

VIEWPOINT



Learning a model,
mean Riemannian covariance (MRC)

OCCLUSION



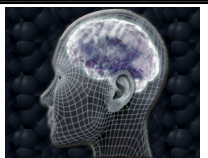
Dense grid of features

DETECTION



Learning a model/MRC discriminants

DISCRIMINATIVE FEATURES



Learning a model/MRC discriminants

(1) Requires annotated data for training

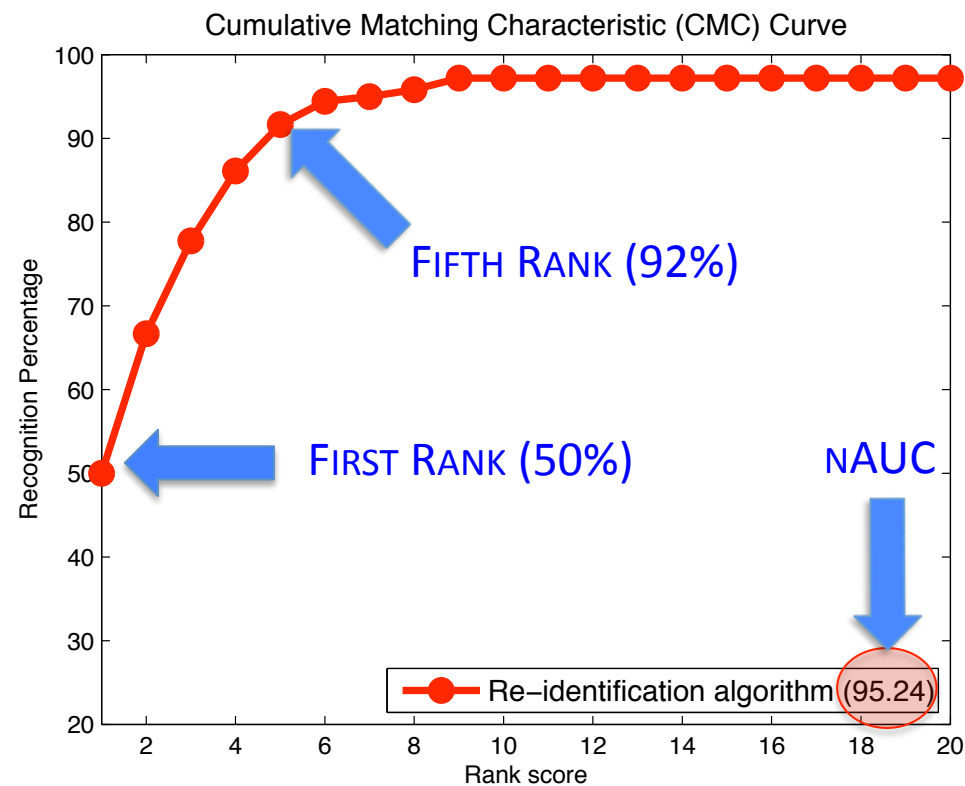


Experimental results

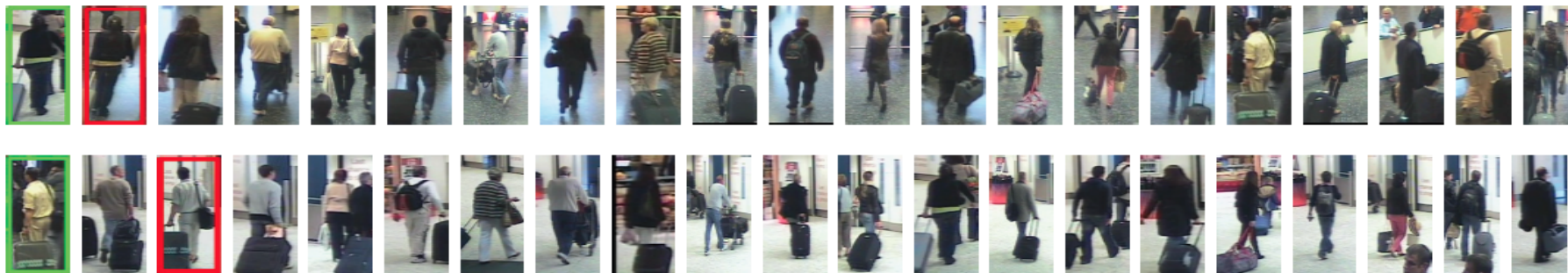
Evaluation metrics

(1) *Cumulative Matching Characteristic curve*

(2) *normalized Area Under Curve (nAUC)* – a quantitative scalar appraisal of CMC



Evaluation scheme based on queries

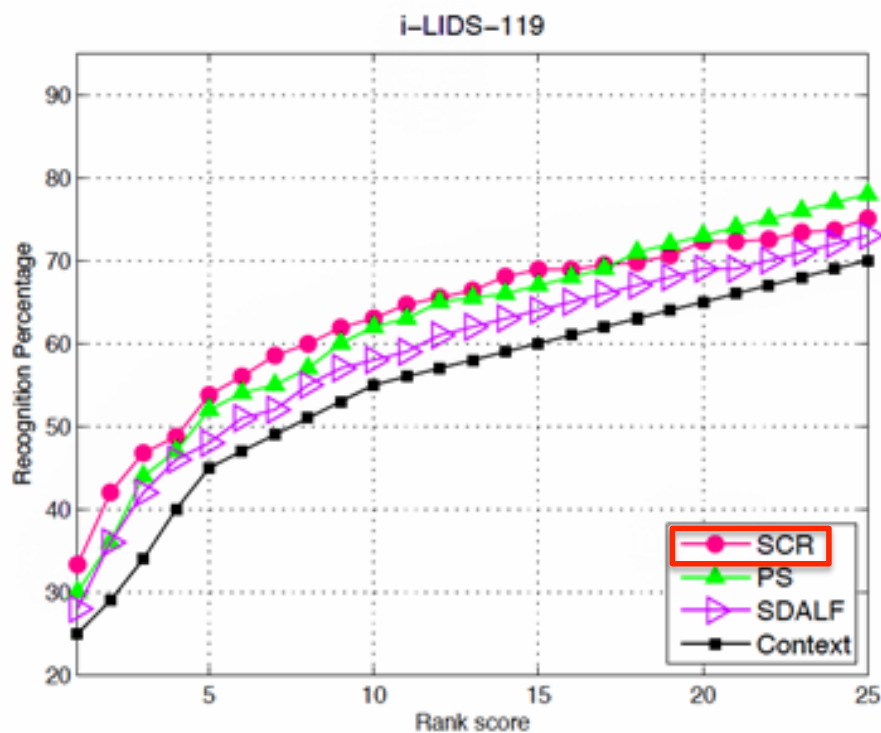




Experimental results -SCR

i-LIDS-119 (Zheng BMVC 2009)

-119 individuals
-in average only 2 images
per person per camera



Single-shot - case

PS [Cheng BMVC 2011]
SDALF [Farenzena CVPR 2010]
Context [Zheng BMVC 2009]



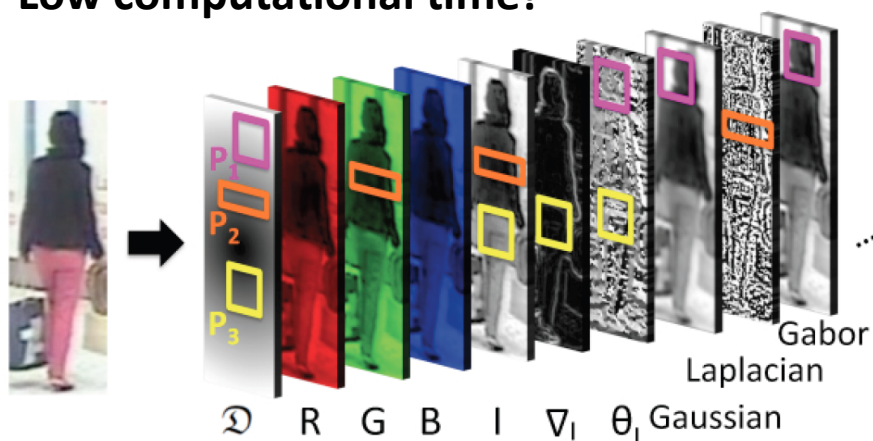
Experimental results -COSMATI

Feature space

Feature layers: color, gradients and filter responses: $L = \{R, G, B, I, \nabla_I, \theta_I, \dots\}$

We select **4x4** covariance matrices: $\mathcal{C} = \{(P, cov(\mathcal{D}, l_i, l_j, l_k)) : l_i, l_j, l_k \in L; P \in \mathcal{P}\}$

Low computational time!



$|\mathcal{P}| = 281$ overlapping rectangular regions (32x8, 16x16 patterns with 8 pixels step)

$$C_{i=1\dots 10} = (R, G, B), (\nabla_R, \nabla_G, \nabla_B), (\theta_R, \theta_G, \theta_B), (I, \nabla_I, \theta_I), (I, G_3, G_4), (I, G_2, \mathcal{L}), \\ (I, G_2, \mathcal{N}), (I, G_1, \mathcal{N}), (I, G_1, \mathcal{L}), (I, G_1, G_2)$$

$$|\mathcal{C}| = 10 \times |\mathcal{P}| = 2810 \xrightarrow{\text{CFS}} 150$$

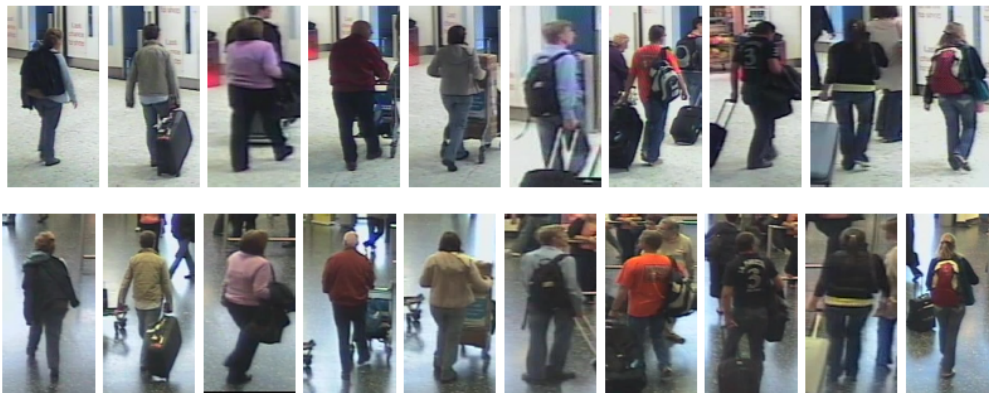


Experimental results -COSMATI

Bak et al, "Boosted human re-identification using Riemannian manifolds", Image and Vision Computing 2011

i-LIDS-MA (our dataset)

- 40 individuals
- 46 images per camera
- manually** detected



i-LIDS-AA (our dataset)

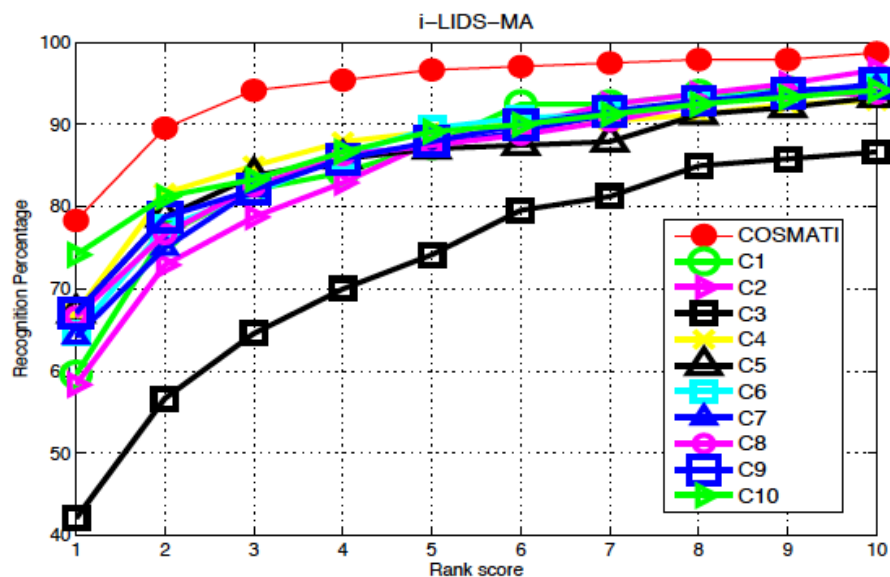
- 100 individuals
- in average 50 images per camera
- automatically** detected



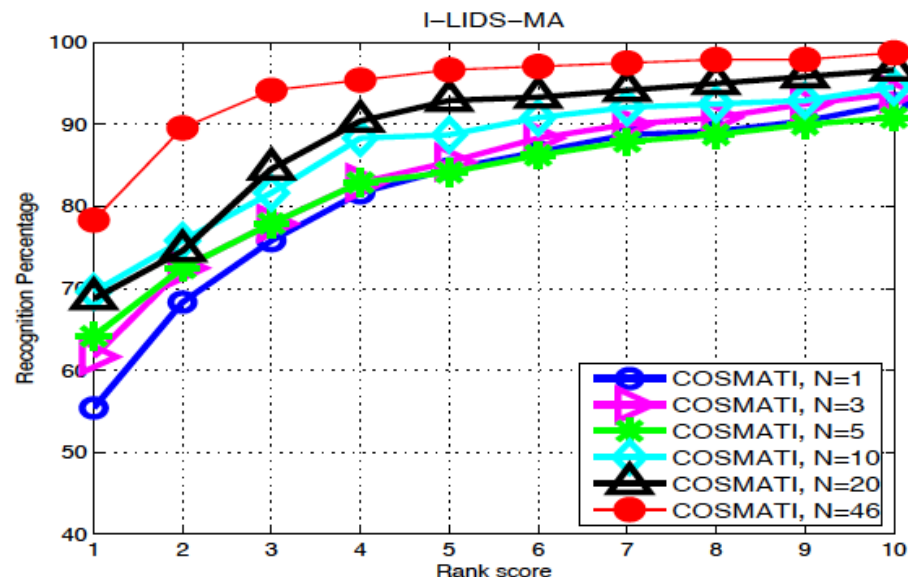


Experimental results -COSMATI

i-LIDS-MA (10 individuals for learning the model and 30 for evaluation)



COSMATI vs single type of covariance

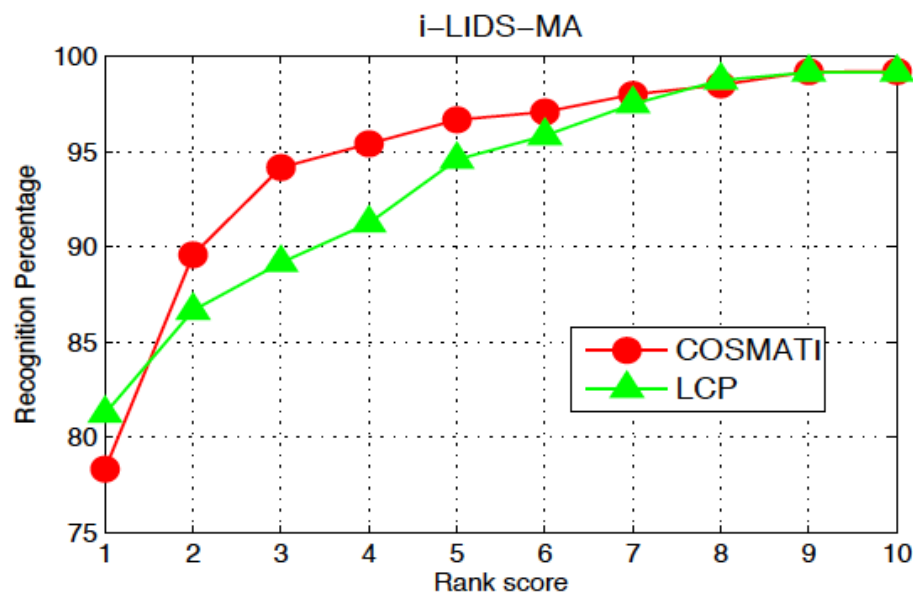


COSMATI w.r.t. the number of given frames

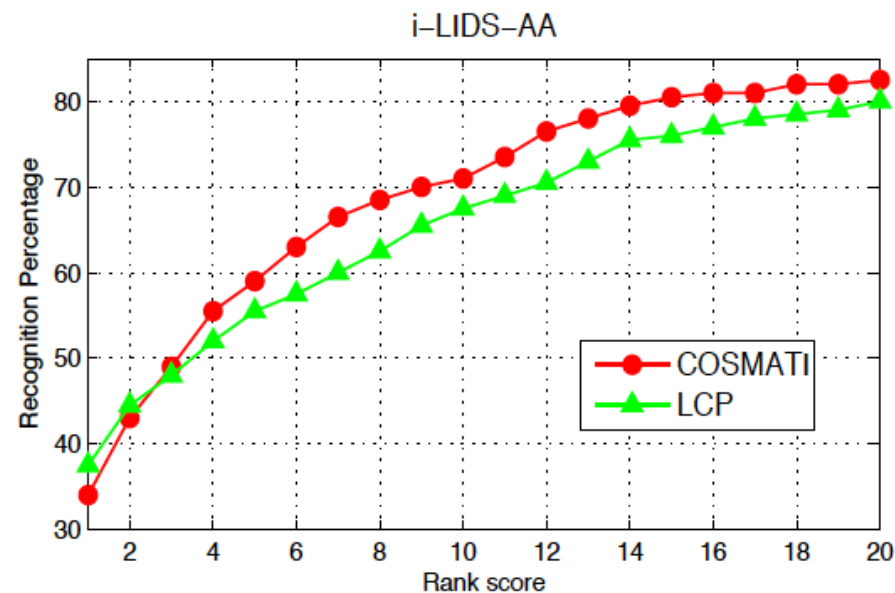


Experimental results -COSMATI

i-LIDS-MA (30 individuals)



i-LIDS-AA (100 individuals)



COSMATI vs LCP (Learned Covariance Patches)

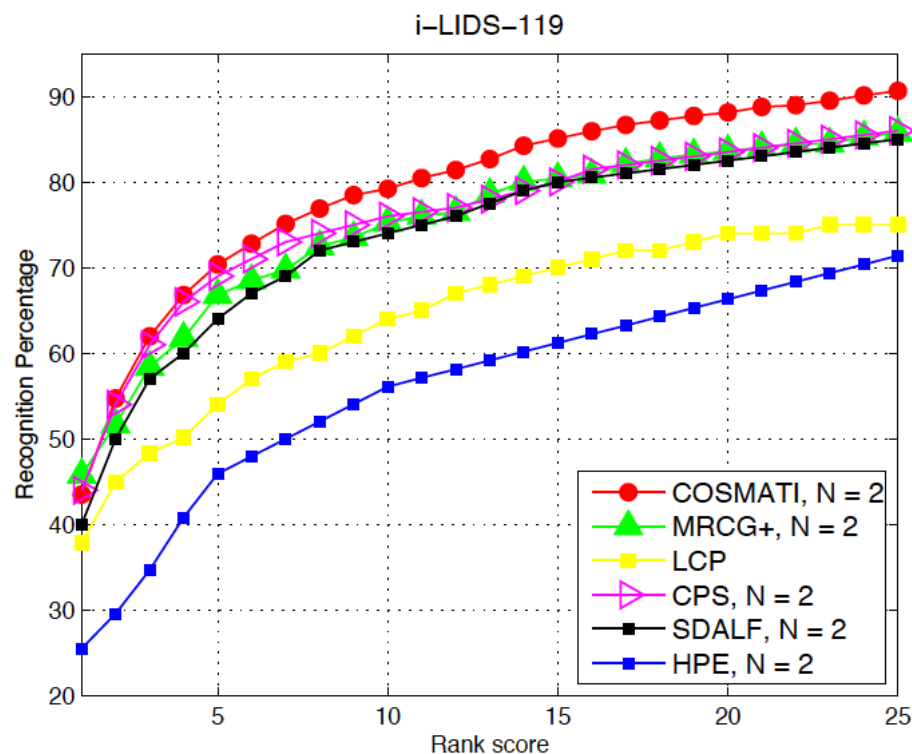
Bak et al, "Boosted human re-identification using Riemannian manifolds", Image and Vision Computing 2011



Experimental results -COSMATI

i-LIDS-119 (Zheng BMVC 2009)

-119 individuals
-in average only 2 images
per person per camera



MRCG+ [Bak AVSS 2011]
LCP [Bak IMAVIS 2011]
CPS [Cheng BMVC 2011]
SDALF [Farenzena CVPR 2010]
HPE [Bazzani ICPR 2010]



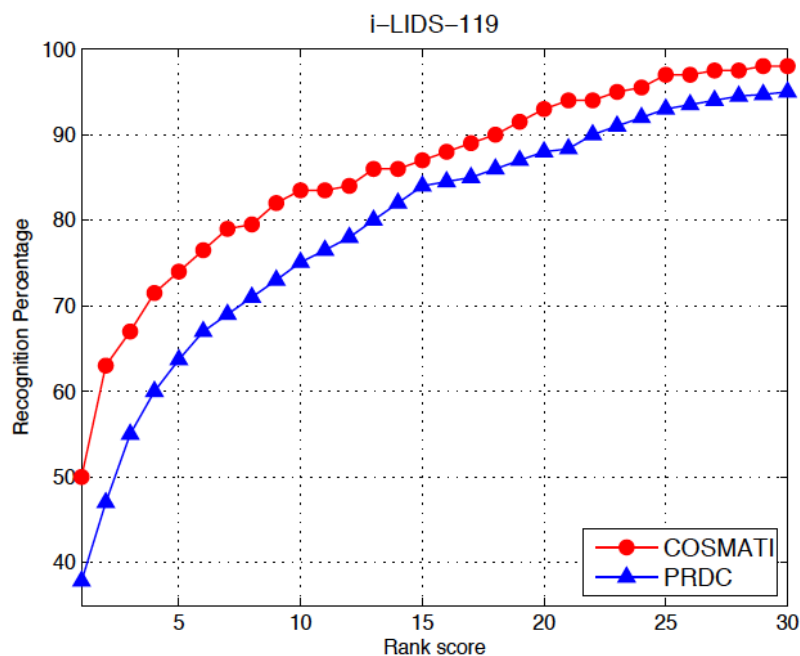
Experimental results -COSMATI

i-LIDS-119

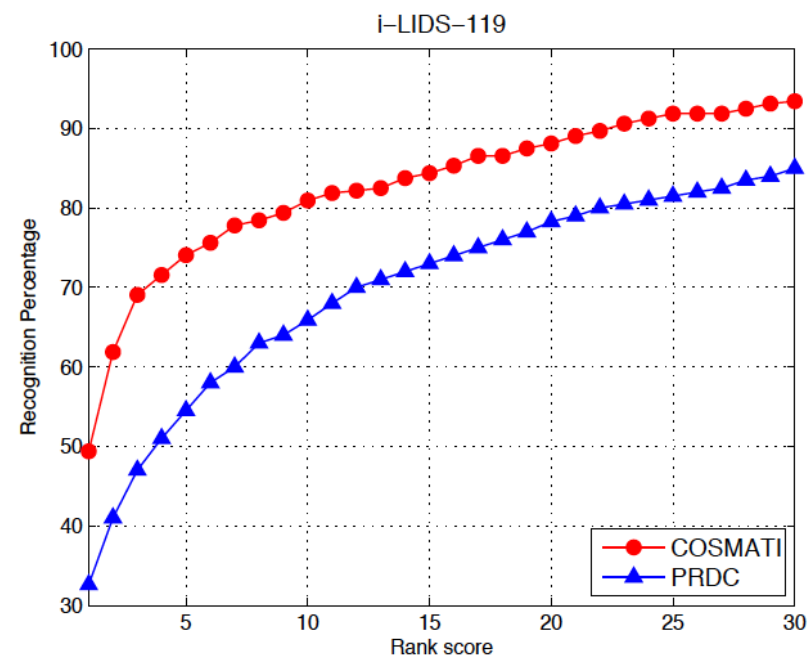
COSMATI vs PRDC [Zheng CVPR 2011]



models extracted on
higher resolution
employ
more texture features



50 individuals



80 individuals



Conclusion

- Propose a **dense grid** of **mean Riemannian covariance** features
 - **color** invariant representation to camera changes,
 - **pose** invariant representation,
 - handled partial **occlusions**,
 - appearance matching problem = **learning** a model
 - Design discriminative approach
 - one-against-all scheme (**MRC discriminants**)
 - **efficient** and **distinctive** representation of object appearance
 - small covariance matrices (**4x4**) between few relevant features
- Use **Body part** driven re-identification + spatial **covariance** features
 - Share **2 new datasets** for evaluation of multiple-shot re-identification

THE END



Thank you for your attention