IS2-10 Meeting on Image Recognition and Understanding (MIRU2016)

Hierarchical Gaussian Descriptor for Person Re-Identification

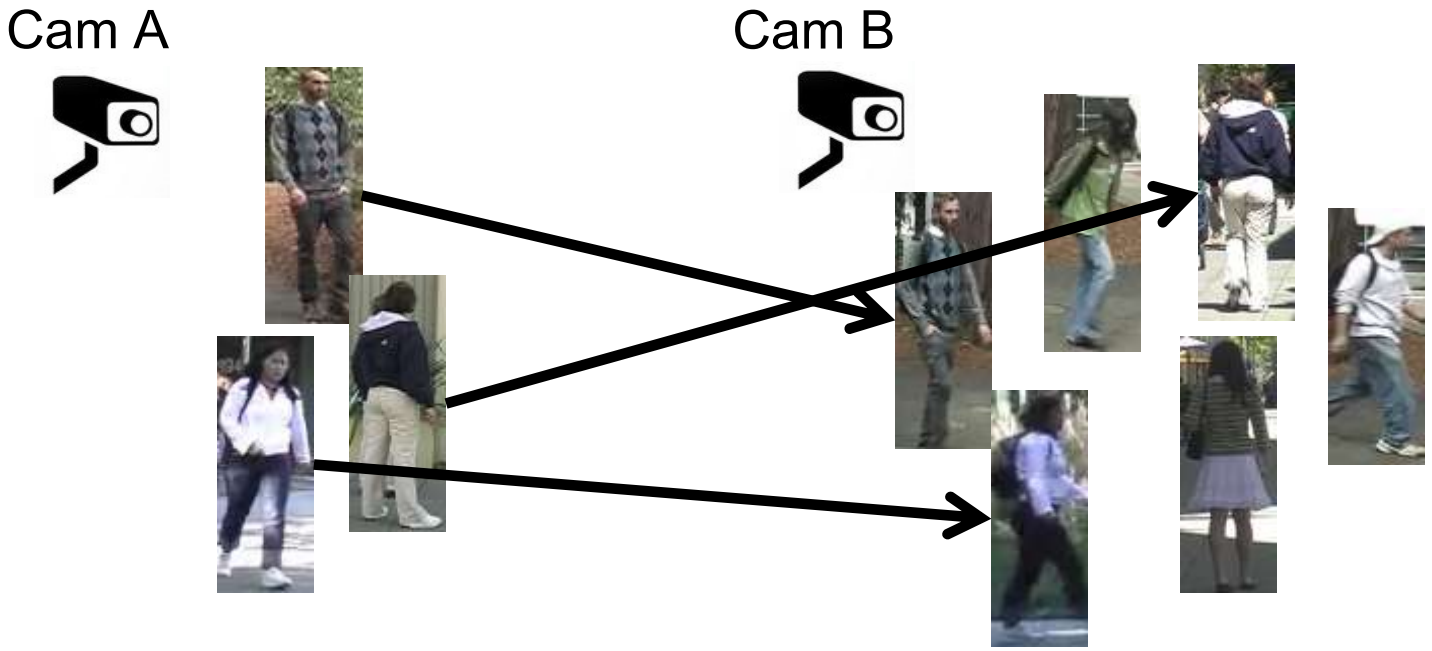
(CVPR2016)

Tetsu Matsukawa[1] Takahiro Okabe[2]
Einoshon Suzuki[1] Yoichi Sato[3]

[1]*Kyushu University,*
[2]*Kyushu Institute of Technology,* [3]*The University of Tokyo*

Person Re-Identification

Search the same person in different camera views

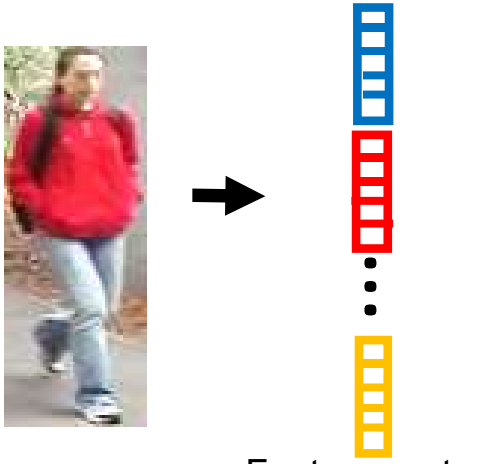


Challenges : large intra-personal variations

e.g. illumination/pose/occlusion/background

Supervised Person Re-Identification

1. Feature extraction



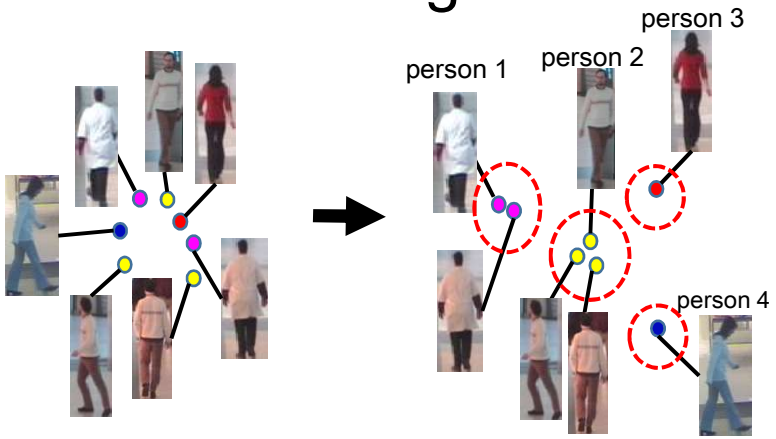
Few works focused on features

- *ELF6* [Gray ECCV2008]
- *Color histograms + LBP* [Hirzer ECCV2012]
- *gBiCov* [Ma IVC2014]
- *LDFV* [Ma ReID2012]
- *SCNCD* [Yang ECCV2014]
- *LOMO* [Liao CVPR2015]

...

This work

2. Metric learning



Many works in the past 5 years

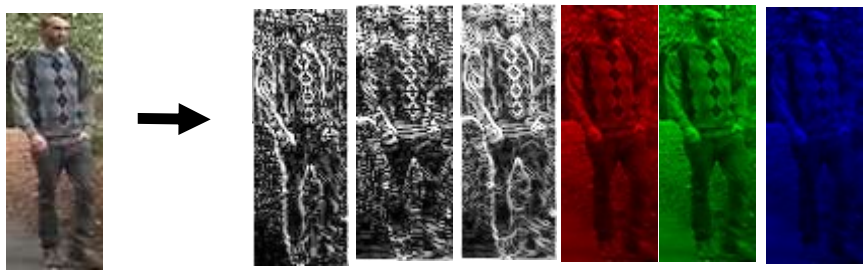
- *LMNN-R* [Dikmen ACCV2010]
- *RDC* [Zheng PAMI2012]
- *RPLM* [Hirzer ECCV2012]
- *KISSME* [Kostinger CVPR2012]
- *LF* [Pedagadi CVPR2013]
- *kLFDA, MFA, rPCCA* [Xiong ECCV2014]
- *XQDA* [Liao CVPR2015]
- *MLAPG* [Liao ICCV2015]

...

Region Covariance Descriptor

[Tuzel ECCV2006]

Pixel feature

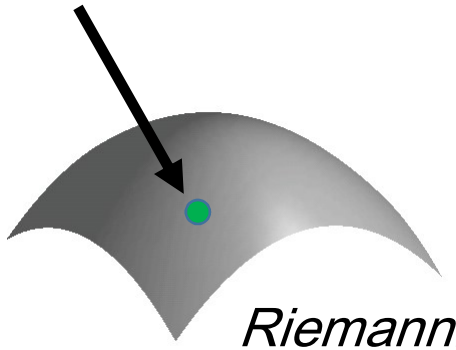


$$\mathbf{f} = \left[\left|I_x\right| \quad \left|I_y\right| \quad \sqrt{I_x^2 + I_y^2} \quad R \quad G \quad B \right]^{\mathrm{T}}$$

Covariance matrix

- *Symmetric Positive Definite* (*SPD*) matrix

$$\boldsymbol{\Sigma} = \frac{1}{N-1} \sum_i \left(\mathbf{f}_i - \bar{\mathbf{f}}\right)\left(\mathbf{f}_i - \bar{\mathbf{f}}\right)^{\mathrm{T}}$$



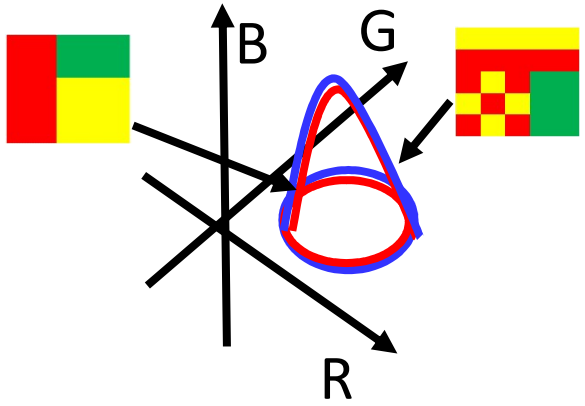
Riemannian manifold

- Natural way to fuse color/texture
- Robust against spatial misalignment
- Applied to person re-identification:

  [Bak AVSS2010, IVC 2012, ECCV2012] [Zeng CVPRWS2015]

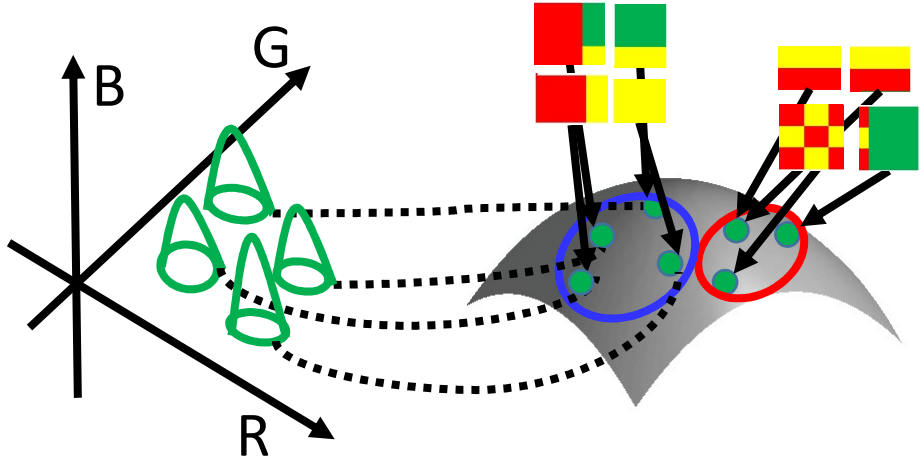
Improvement 1:
Hierarchical distribution

Global distribution (*conventional*)



- Less discriminative
  (loses local structure)

Distribution on local distributions (*this work*)



- More discriminative
  (includes local structure)

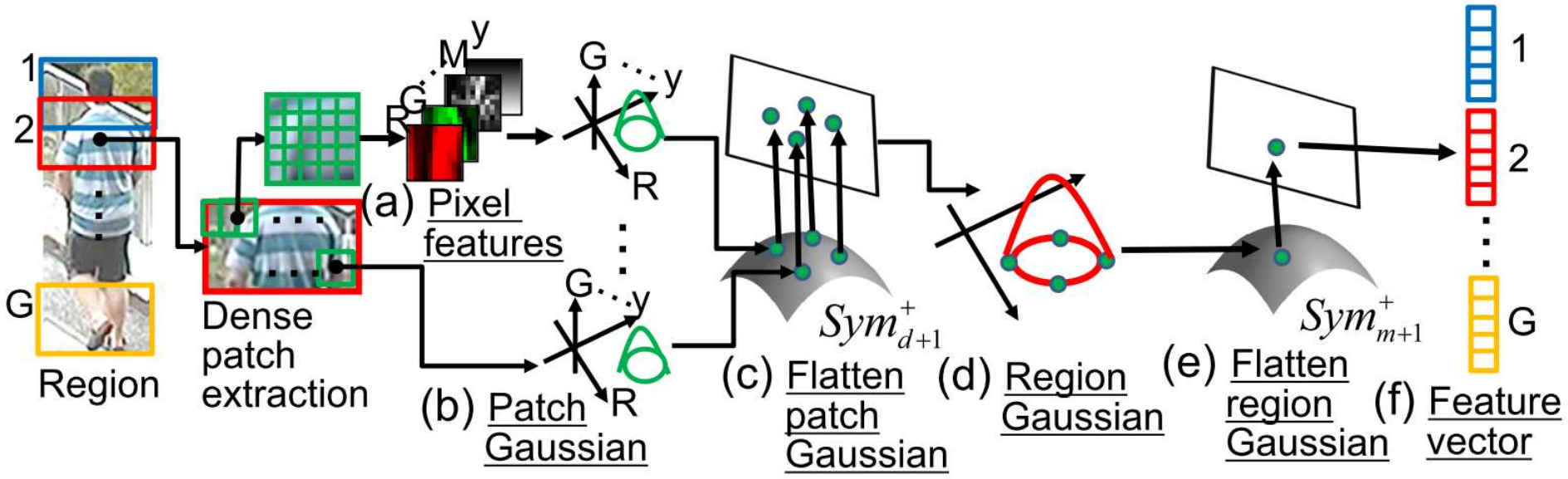
Improvement 2:
Mean color of local region

- Major discriminative information is mean color (absent in covariance)

This work proposes a hierarchical Gaussian descriptor

- High accuracy: includes both mean and covariance in patch/region levels
- No learning process: hand crafted descriptor

Gaussian of Gaussians (GOG) Descriptor

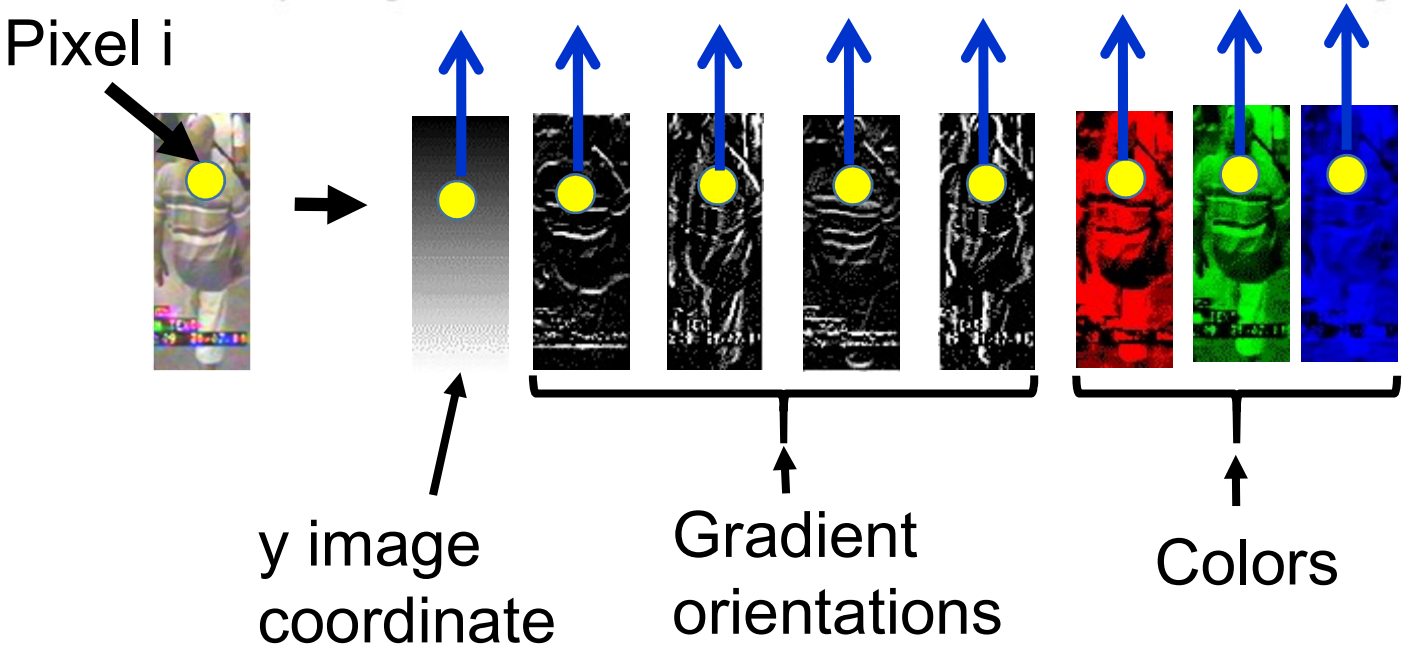


(a) Pixel Features

- Characterize each pixel by d dim. feature vector

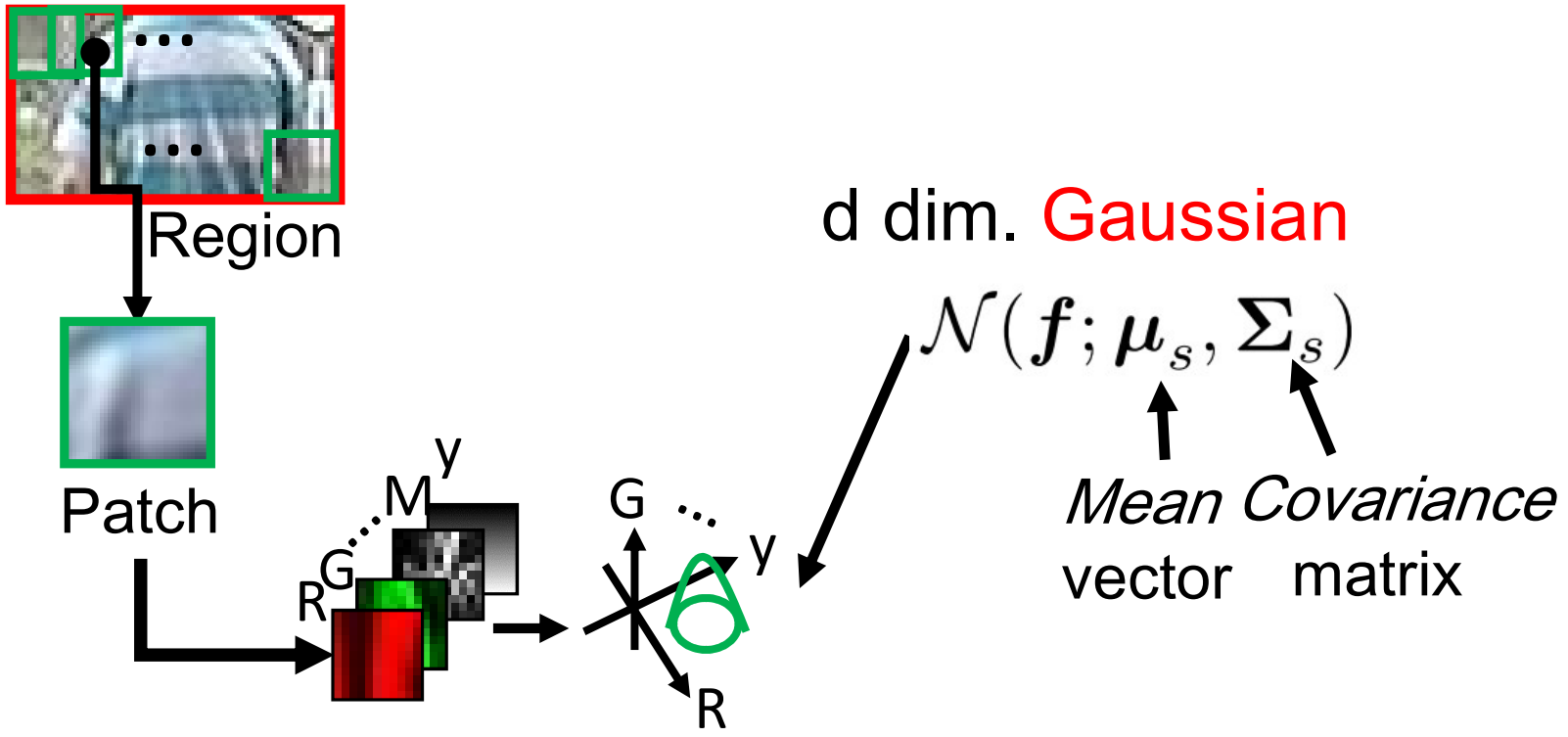
e.g. d = 8

$$\boldsymbol{f}_i = [y, M_{0^\circ}, M_{90^\circ}, M_{180^\circ}, M_{270^\circ}, R, G, B]^T$$



(b) Patch Gaussians

- Summarize pixel features inside a patch



(c) Flatten Patch Gaussians

- Vectorization by considering underlying *Riemannian geometry*

1. Embed Gaussians into *SPD* matrix manifold (Sym^+_{d+1})
   [Lovric et al. 2000]

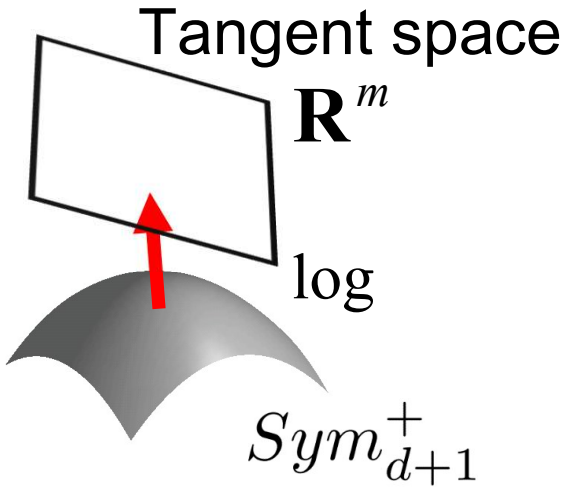
$$\mathcal{N}(\boldsymbol{f};\boldsymbol{\mu}_s,\boldsymbol{\Sigma}_s) \sim \boldsymbol{P}_s = |\boldsymbol{\Sigma}_s|^{-\frac{1}{d+1}} \begin{bmatrix} \boldsymbol{\Sigma}_s + \boldsymbol{\mu}_s\boldsymbol{\mu}_s^T & \boldsymbol{\mu}_s \\ \boldsymbol{\mu}_s^T & 1 \end{bmatrix}$$

2. Apply log Euclidean[Arsigny et al. 2006] and half vectorization

$$\boldsymbol{g}_s = \mathrm{vec}(\log(\boldsymbol{P}_s))$$

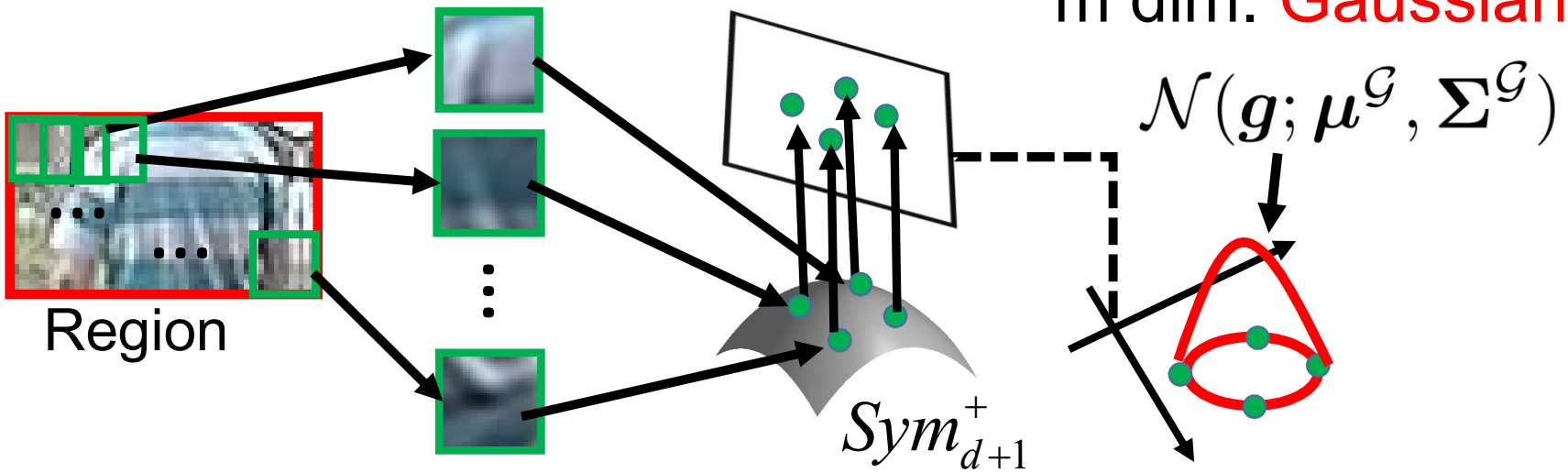
m = $(d^2+3d)/2 + 1$ dim.

e.g. m = 45 when d = 8

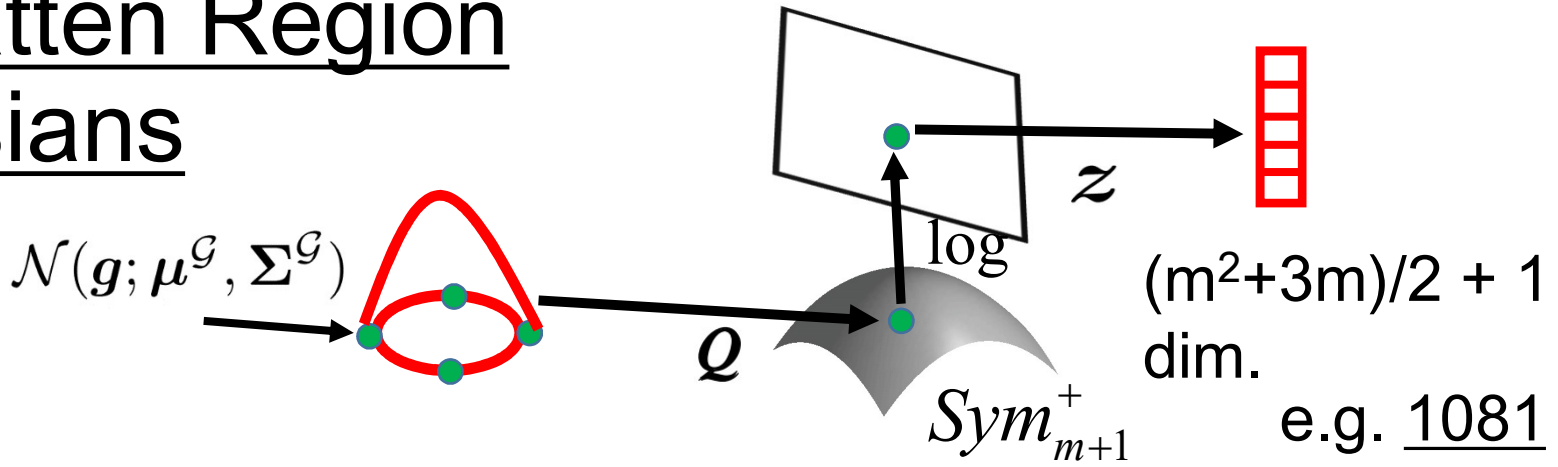


(d) Region Gaussian

- Summarize patch Gaussians inside a region



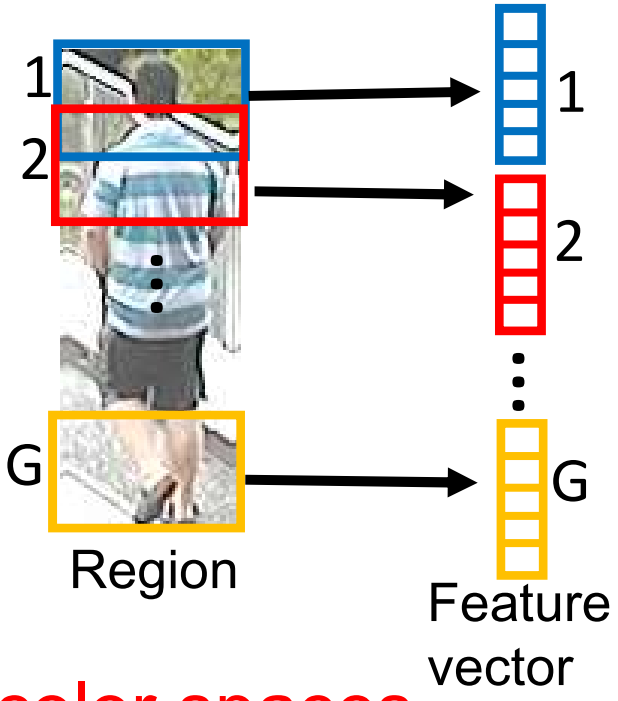
(e) Flatten Region Gaussians



(f) Feature Vector

- Concatenate feature vectors of G regions

  e.g. 7 region x 1081 dim. = 7567 dim.



Color Space Fusion

- Concatenate GOGs of different color spaces

  e.g. 3 spaces x 7567 dim. + 1 space x 4921 dim.

  (RGB, Lab, HSV) (nRnG)

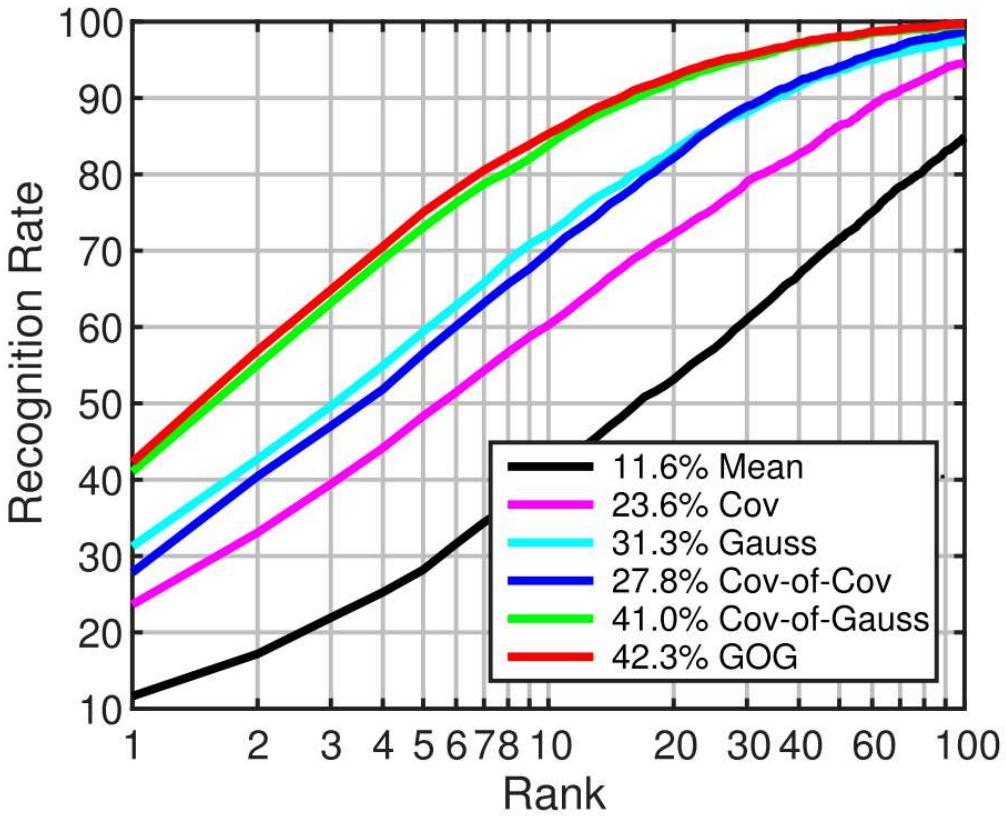
  = 27,622 dim.

Experimental Setting

- Metric learning : *Cross-view Quadratic Discriminant Analysis (XQDA)* [Liao CVPR2015]
- On five public datasets

VIPeR PRID450S GRID CUHK01&03

Performance Analysis on VIPeR



- + Hierarchy
  (Gauss → GOG)
  improves 11.0% (rank-1)

- + Mean
  (Cov-of-Cov → GOG)
  improves 14.5% (rank-1)

Comparison with State-of-the-Art

Rank-1 rates

Methods	Ref.	VIPeR	CUHK01	PRID450S	GRID
GOG_{Fusion}+XQDA	Ours	49.7	57.8	68.8	24.7
Metric Ensemble	CVPR2015	45.9	53.4	-	-
LOMO+XQDA	CVPR2015	40.0	49.2	62.6	16.6
SCNCD	ECCV2014	37.8	-	41.6	-
Semantic	CVPR2015	31.1	32.7	43.1	-
SalMatch	ICCV2013	30.2	28.5	-	-
MLFL	CVPR2014	29.1	34.3	-	-

On Automatic Detected Dataset

Rank-1 rates of CUHK03

Methods	Ref.	Labeled	Detected
GOG_{Fusion}+XQDA	Ours	67.3	65.5
Metric Ensemble	CVPR2015	62.1	-
LOMO+MLAPG	ICCV2015	58.0	51.2
LOMO+XQDA	CVPR2015	52.2	46.3
ImprovedDeep	CVPR2015	54.3	45.0
DeepReID	CVPR2014	20.7	19.9

- Performance drop between manual bounding box and automatic bounding box is 1.8% (rank-1 rate).

Example results (CUHK01)

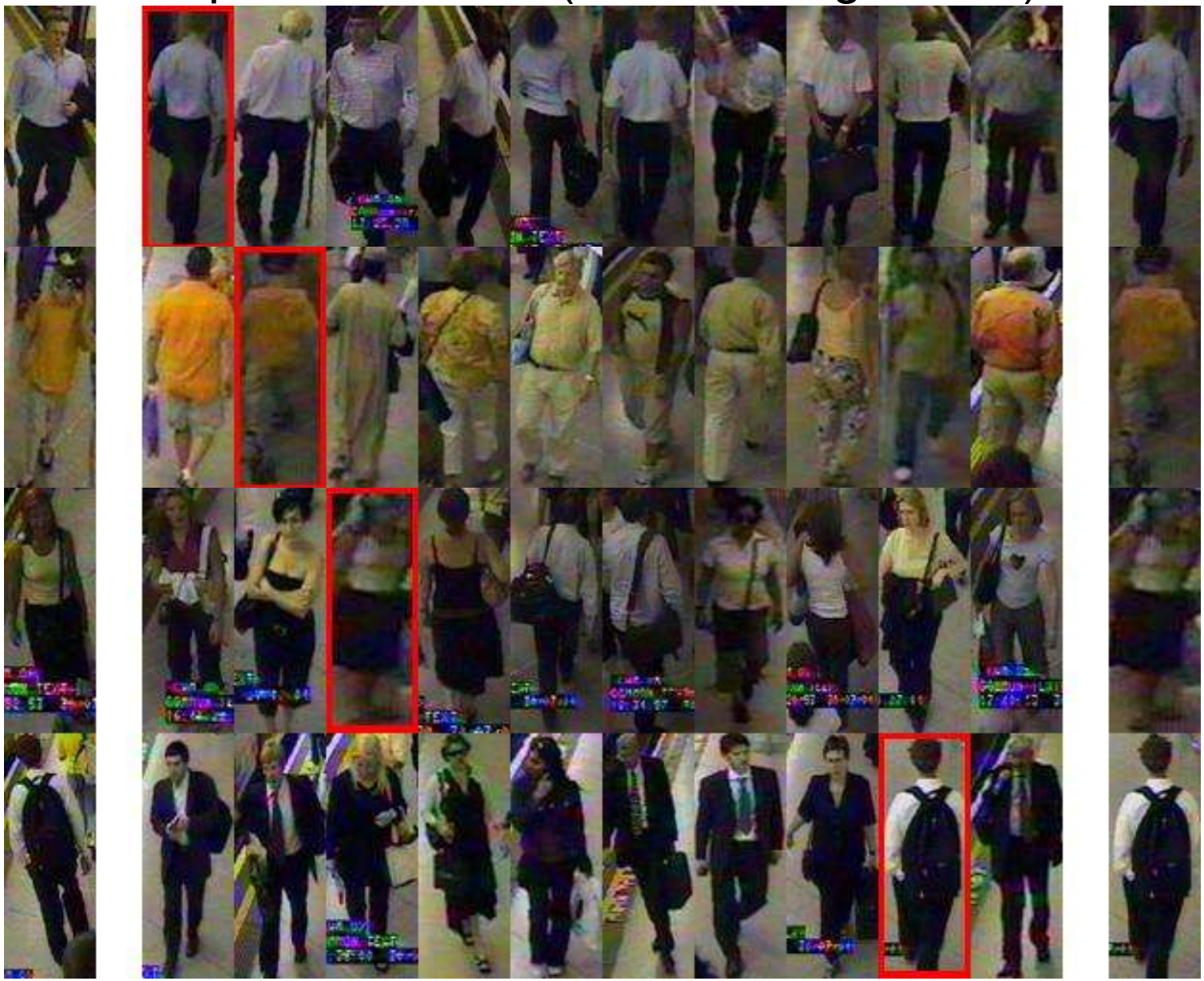
Probe Top 10 candidate (from 486 galleries) Ground truth

Example results (CUHK01)

Probe Top 10 candidate (from 486 galleries) Ground truth

Example results (GRID)

Probe | Top 10 candidate (from 1025 galleries) | Ground truth



Conclusion

- Proposed a novel *hierarchical (two-level) Gaussian* descriptor

- High performance which improves the state-of-the-art performances on five public datasets

- Matlab code is available:
http://www.i.kyushu-u.ac.jp/~matsukawa/ReID.html