Hierarchical Gaussian Descriptor for Person Re-Identification

Tetsu Matsukawa[1], Takahiro Okabe[2], Einoshin Suzuki[1], Yoichi Sato[3]
[1]*Kyushu University*, [2]*Kyushu Institute of Technology*, [3]*The University of Tokyo*

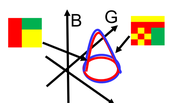
IEEE 2016 Conference on Computer Vision and Pattern Recognition
CVPR 2016

Motivation

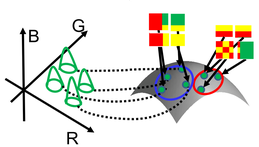
Covariance descriptor

- Natural way to fuse color/texture
- Robust against spatial misalignment

Hierarchical distribution

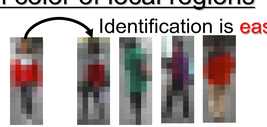


- Global distribution is less discriminative (loses local structure)



- Distribution on local distributions is more discriminative (includes local structure)

Mean color of local regions



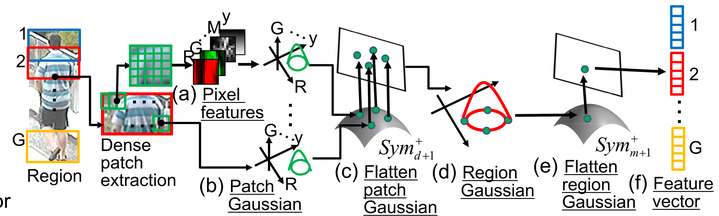
- Major discriminative information is mean color (absent in covariance)

Related descriptors

	Covariance	Gaussian
Global	[Tuzel06]	[Gong09][Nakayama10] [Ma14][Serra15]
Hierarchical (two level)	[Li12] [Serra14]	[Ours]

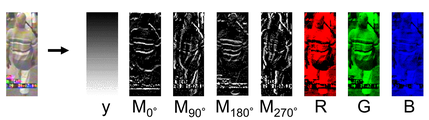
Gaussian of Gaussians (GOG) descriptor

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



(a) Pixel features

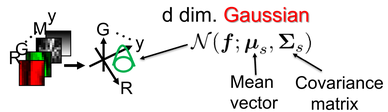
- Characterize each pixel by d dim. feature vector $\boldsymbol{f}$



e.g. d = 8

(b) Patch Gaussian

- Summarize pixel features inside a patch



(c) Flatten patch Gaussian

- Flatten patch Gaussians by considering underlying *Riemannian Geometry* of Gaussians

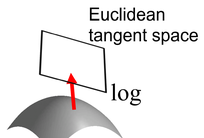
1. Embed Gaussians into *Symmetric Positive Definite (SPD) matrix* manifold (Sym^{+}_{d+1}) [Lovric00]

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

2. Apply log Euclidean [Arsigny06] and half vectorization

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

m = (d²+3d)/2 + 1 dim.
e.g. m = 45 when d = 8



(d) Region Gaussian

- Summarize patch Gaussians inside a region in m dim. Gaussian: $\mathcal{N}(\boldsymbol{g};\boldsymbol{\mu}^{\mathcal{G}},\boldsymbol{\Sigma}^{\mathcal{G}})$
- Weighted mean/covariance is used

$$\boldsymbol{\mu}^{\mathcal{G}} = \frac{1}{\sum_{s\in\mathcal{G}} w_s} \sum_{s\in\mathcal{G}} w_s \boldsymbol{g}_s$$

$$\boldsymbol{\Sigma}^{\mathcal{G}} = \frac{1}{\sum_{s\in\mathcal{G}} w_s} \sum_{s\in\mathcal{G}} w_s (\boldsymbol{g}_s - \boldsymbol{\mu}^{\mathcal{G}})(\boldsymbol{g}_s - \boldsymbol{\mu}^{\mathcal{G}})^T$$

(e) Flatten region Gaussians

1. Embed region Gaussian into Sym^{+}_{m+1}: $\mathcal{N}(\boldsymbol{g};\boldsymbol{\mu}^{\mathcal{G}},\boldsymbol{\Sigma}^{\mathcal{G}}) \sim \boldsymbol{Q}$
2. Apply log Euclidean and half vectorization: $\boldsymbol{z} = \mathrm{vec}(\log(\boldsymbol{Q}))$

(m²+3m)/2 + 1 dim.
e.g. 1081 when m = 45

(f) Feature vector

- Concatenate feature vectors of G regions
  e.g. 7 region x 1081 dim. = 7567 dim.

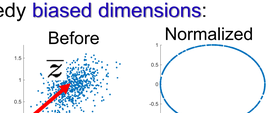
Color space fusion

- Concatenate GOGs of different color spaces
  e.g. 3 spaces x 7567 dim. + 1 space x 4921 dim. = 27,622 dim.
  (RGB, HSV, YCbCr) (nRnG)

Normalization

- Sample mean removal is required before performing L2 norm normalization to remedy biased dimensions:

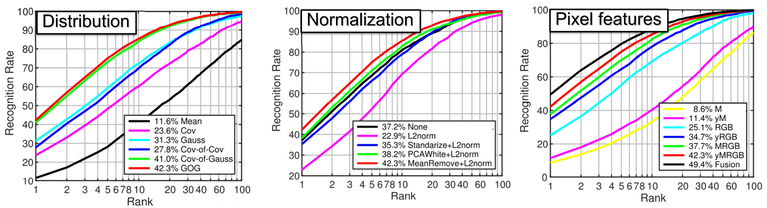
$$\boldsymbol{z} = (\boldsymbol{z} - \overline{\boldsymbol{z}})/\|\boldsymbol{z} - \overline{\boldsymbol{z}}\|_2$$



Experimental Results

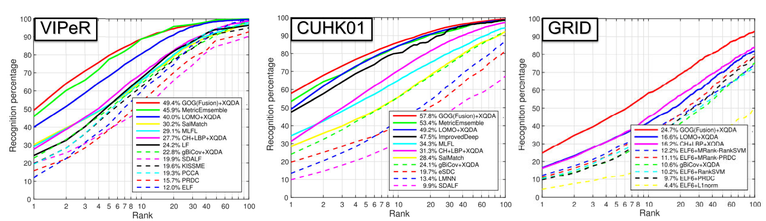
Performance analysis on VIPeR

- XQDA metric learning [Liao15] is applied to GOG



- + Hierarchy (Gauss → GOG) improves 11.0% (rank-1)
- + Mean (Cov-of-Cov → GOG) improves 14.5% (rank-1)
- L2 norm degrades 14.3% (rank-1)
- Mean removal + L2 norm improves 5.1% (rank-1)
- Adding pixel features improves performance
- Color fusion improves 7.1% (rank-1)

Performance comparison



- GOG + XQDA achieves the new state-of-the art 49.7%, 57.8%, 24.7% rank-1rate on VIPeR, CUHK01, GRID dataset

On automatic detected dataset

CUHK03 (rank-1 rate)

Methods	Labeled	Detected
GOG_{Fusion}+XQDA	67.3	65.5
Metric Ensemble	62.1	-
LOMO+XQDA	52.2	46.3
Improved Deep	54.3	45.0
Deep ReID	20.7	19.9

Feature extraction time

Seconds / image

Methods	48x128 pixel
LOMO	0.016
Cov_{RGB}	0.021
GOG_{RGB}	0.34
GOG_{Fusion}	1.34
gBiCov	7.8

Project page : http://www.i.kyushu-u.ac.jp/~matsukawa/ReID/