# Person Re-Identification via Discriminative Accumulation of Local Features

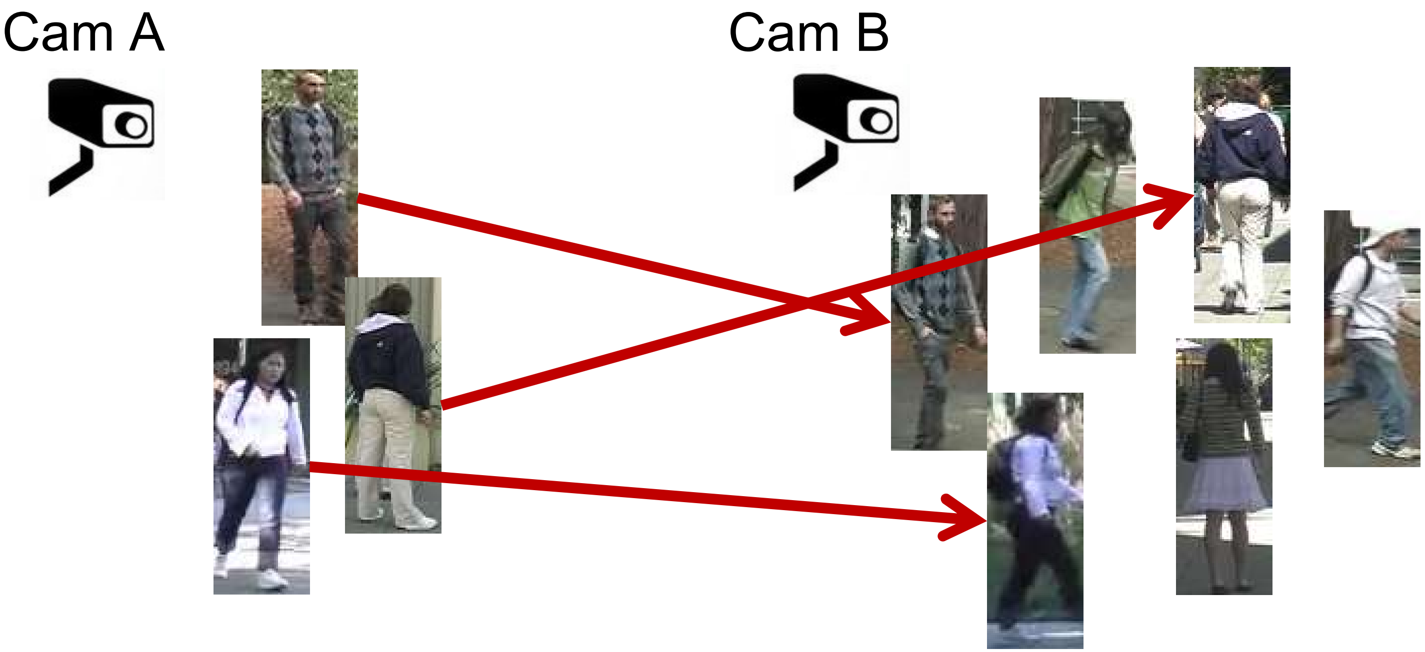
Tetsu Matsukawa[1] Takahiro Okabe[2]
Yoichi Sato[1]

[1]The University of Tokyo
[2]Kyushu Institute of Technology

22nd INTERNATIONAL
CONFERENCE ON PATTERN
RECOGNITION
2014

# Person Re-Identification

- Search same person in different camera



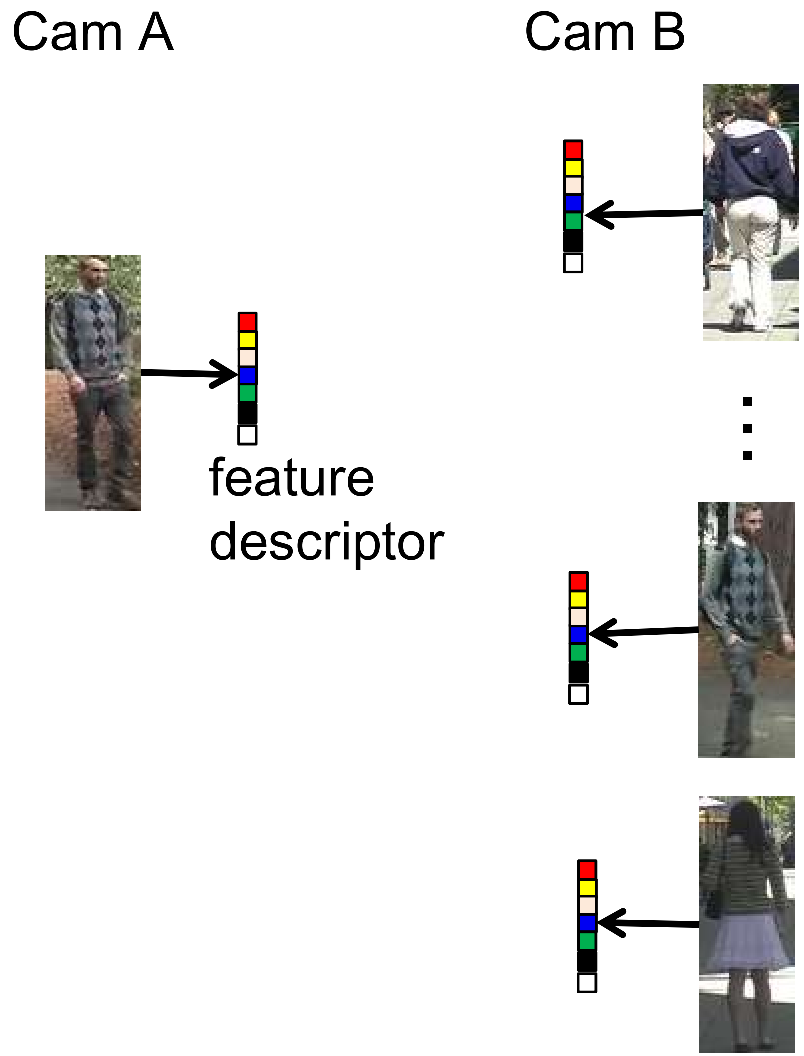

Applications suspected person search/trajectory analysis

Challenges large intra-personal variations

e.g. illumination/pose/occlusion/background

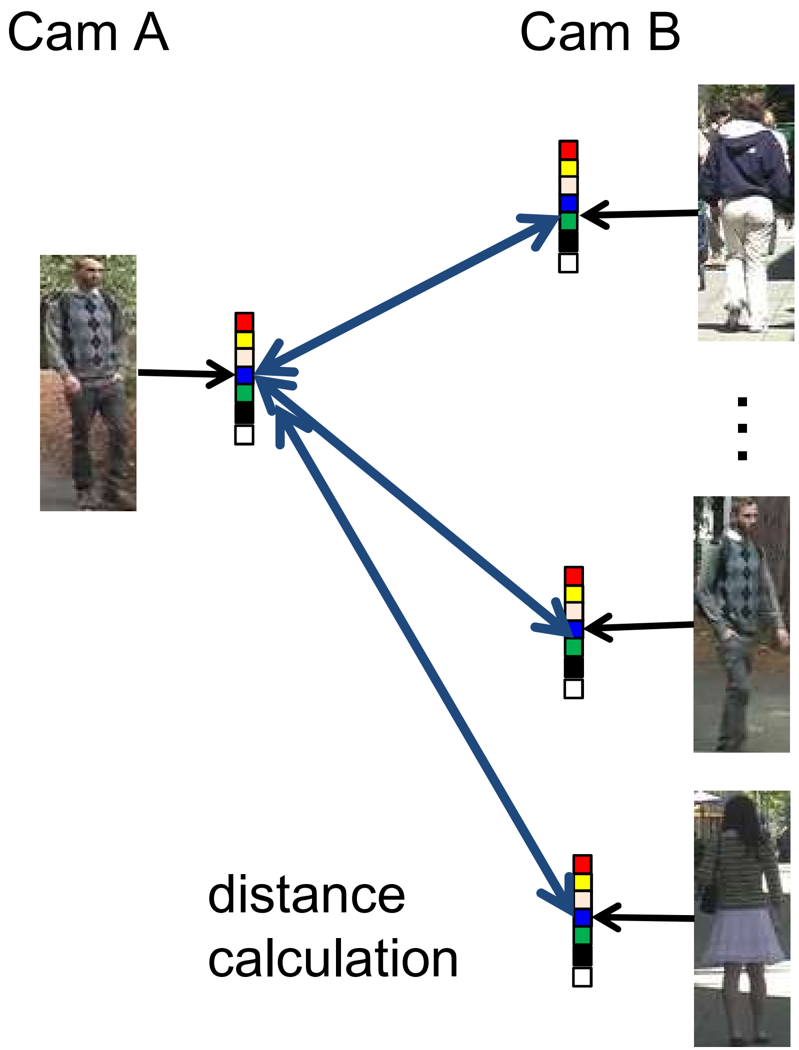
# Two Main Steps

- Extract feature descriptor



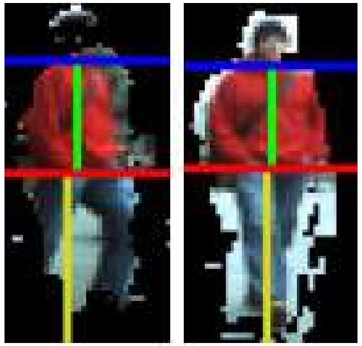
# Two Main Steps

- Extract feature descriptor
- Match feature descriptor



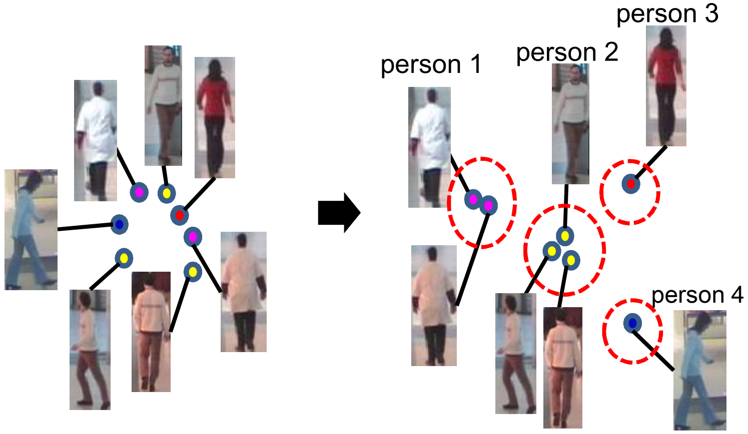
# Related Work of Re-Identification

- Extract feature descriptor
  - Distinctive but robust to pose variation
    - *Dense color histograms* [Hirzer12]
    - *Symmetric driven accumulation* [Bazzani13]
    - *Hierarchical gaussian map* [Hu13]
    - …



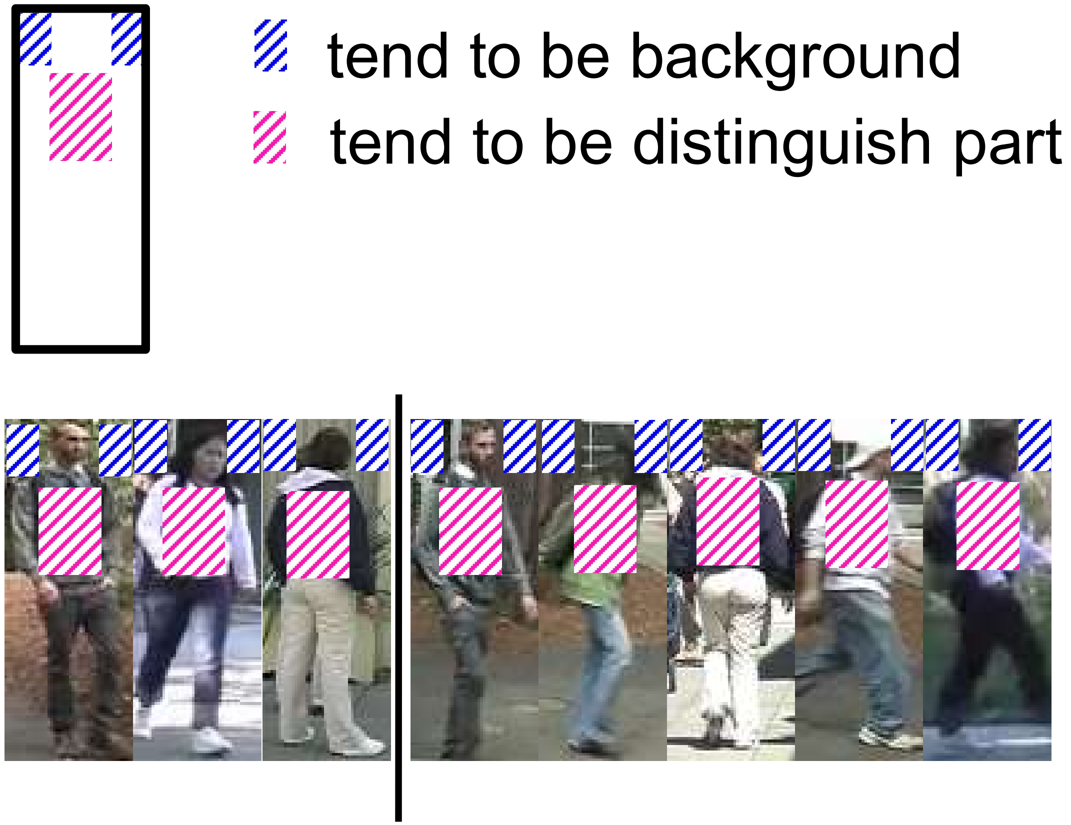
Bazzani13

- Match feature descriptor
  - Supervised learning of distance metric
    - *LMNN* [Dikmen10]
    - *RDC* [Zheng12]
    - *RPLM* [Hirzer12]
    - *KISSME* [Kostinger12]
    - *LF* [Pedagadi13]
    - …

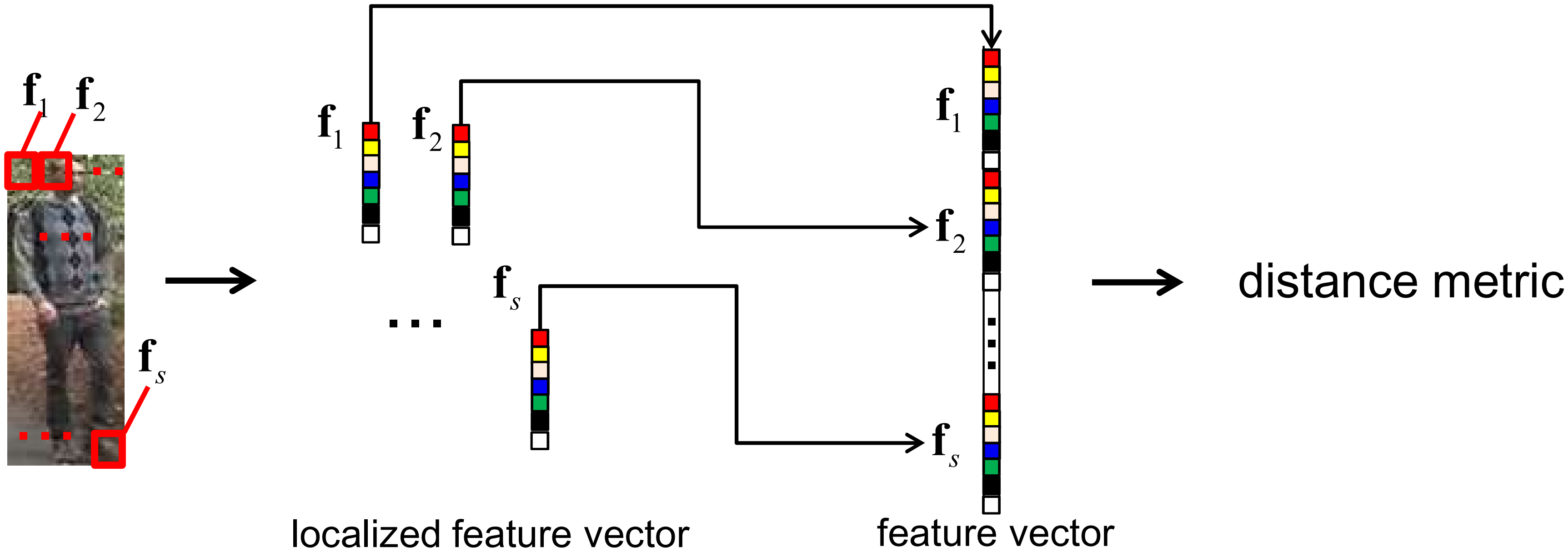


# Key Observation

- There exists discriminative positions commonly on dataset

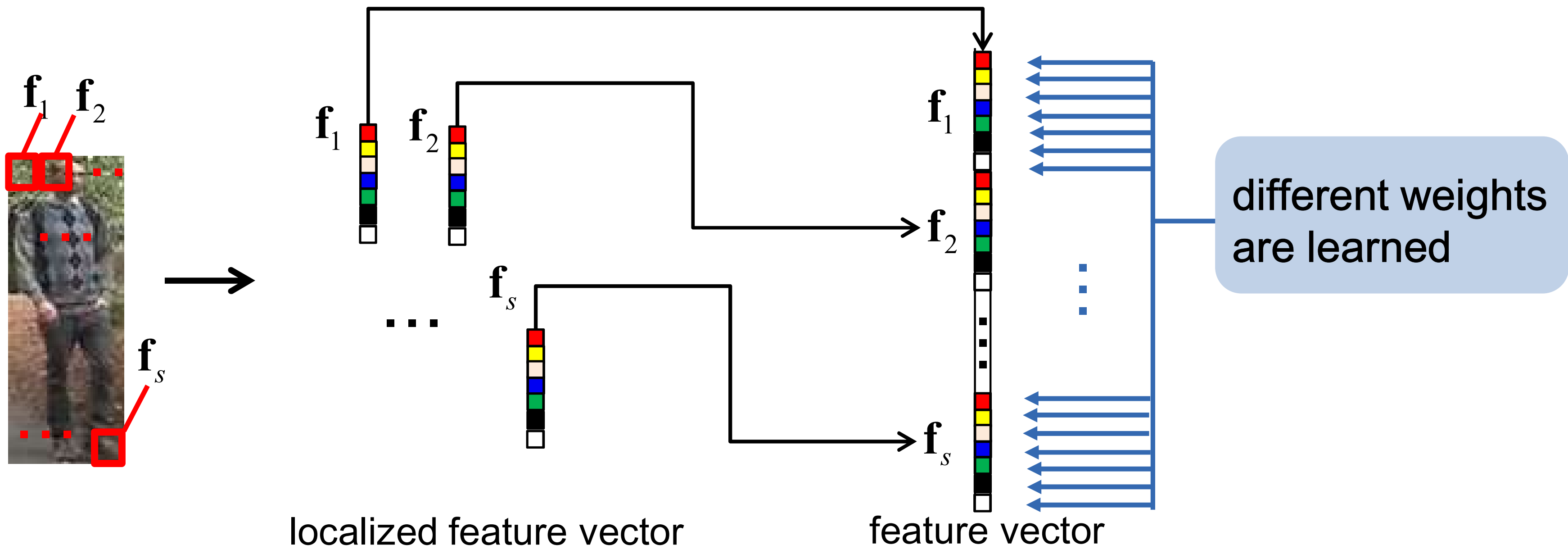


# Conventional Feature Form




- Concatenated vector is used

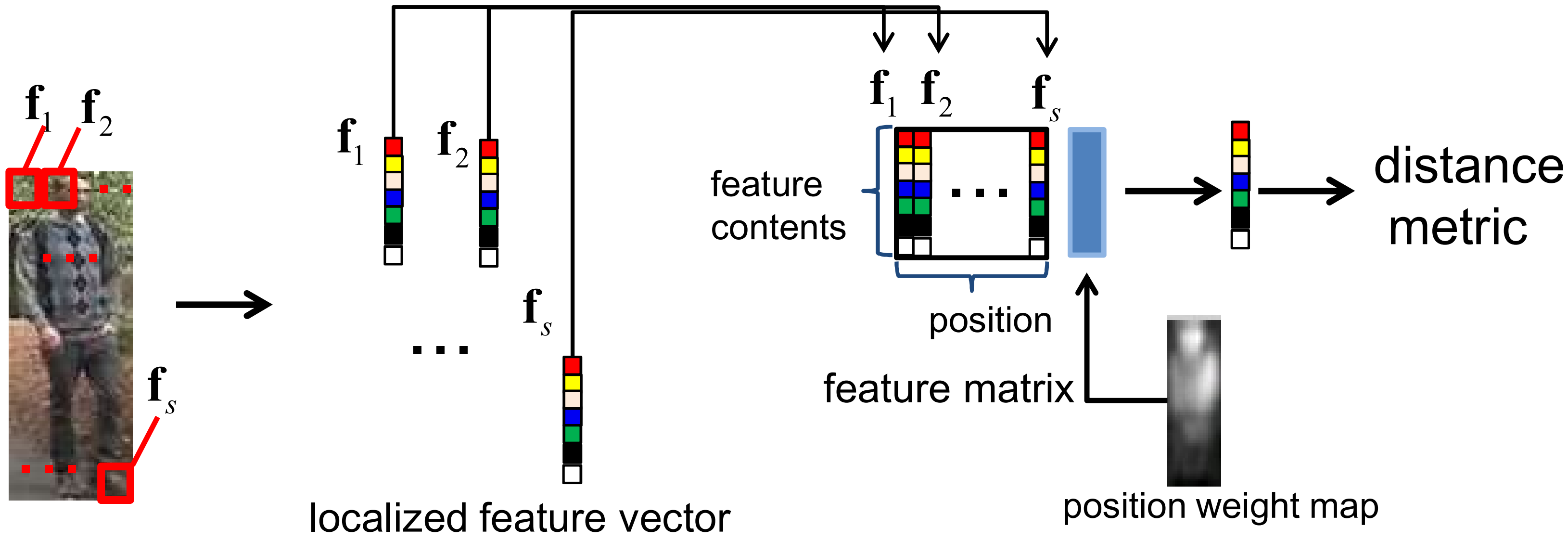
# Conventional Feature Form




- Concatenated vector is used
  - The discriminative positions are separately analyzed for each of feature dimensions (e.g. color histogram bin)

- When the number of training samples is not enough --- over-fitting

# Our Main Idea

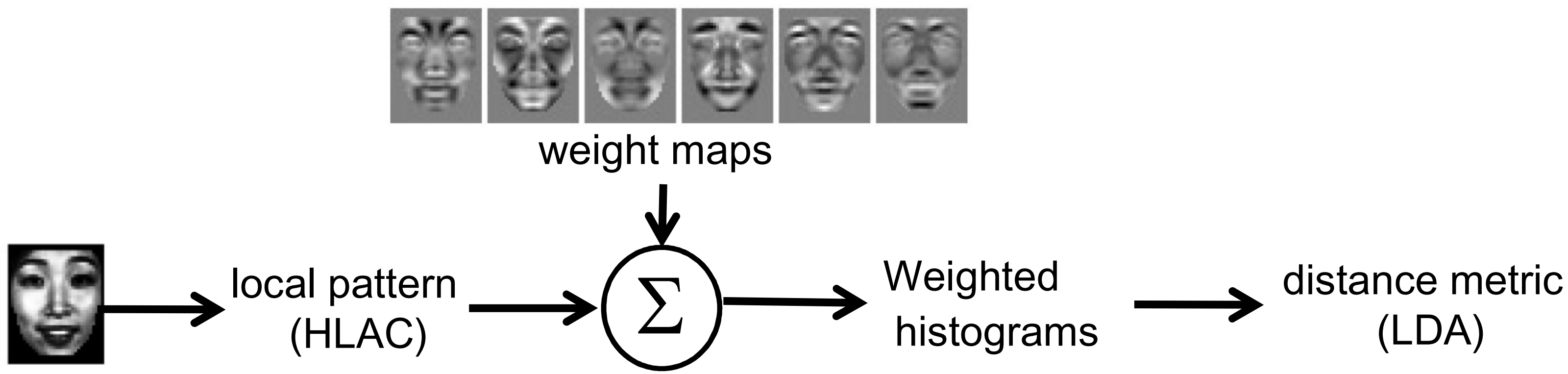



- Construct feature matrix
  - Analyze shared position importance for all feature dimensions

- The parameter to be learned is small ---- less over-fitting

# Related Work

- **Fisher Weight Map (FWM )** [Shinohara04]
  - Discriminatively learned weight maps for weighted histogram for facial expression recognition



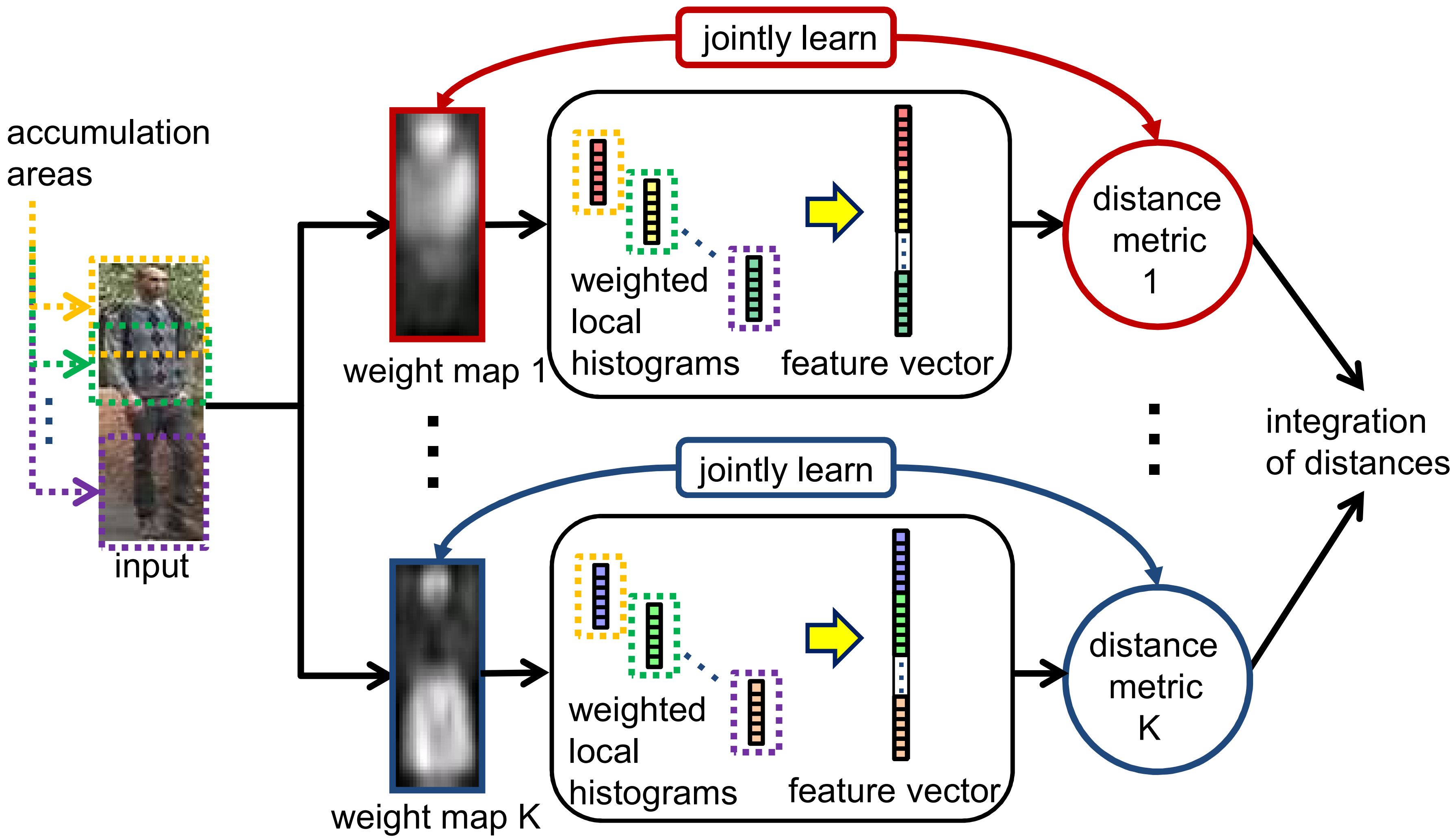

**FWM**

- Global histogram
- Separately learn weight map and distance metric

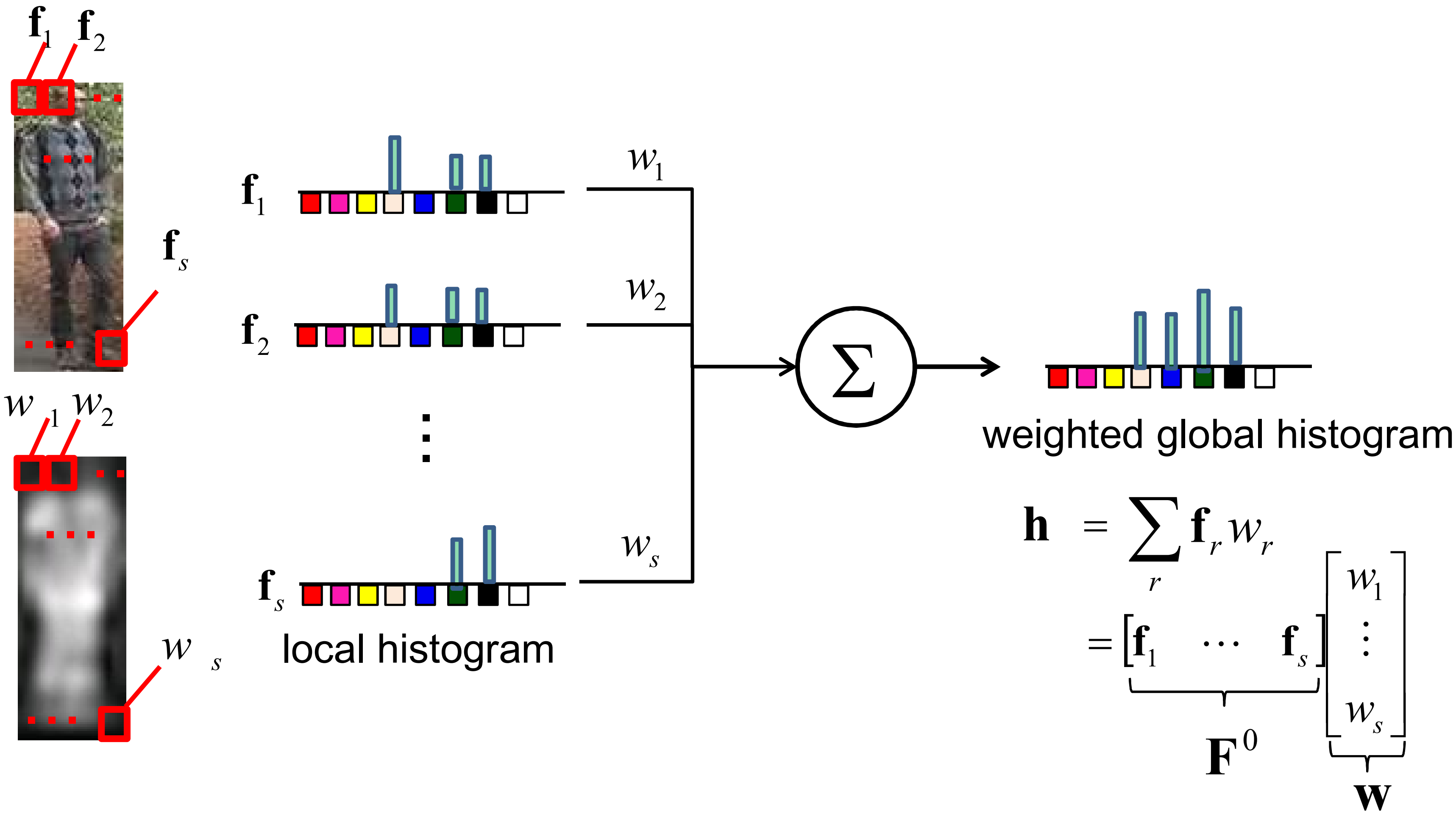
**Ours**

- Local histogram of wide area
- Jointly learn weight map and distance metric

# Flow of the Proposed Method



# Linear Form of a Weighted Histogram [Shinohara04]

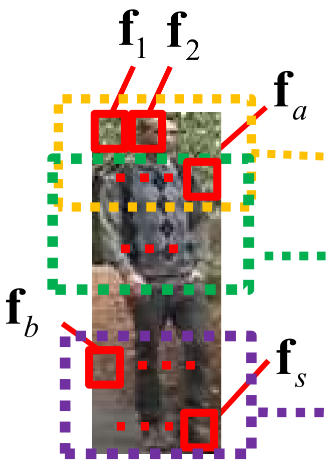



$$\mathbf{h} = \sum_r \mathbf{f}_r w_r = \underbrace{[\mathbf{f}_1 \ \cdots \ \mathbf{f}_s]}_{\mathbf{F}^0} \underbrace{\begin{bmatrix} w_1 \\ \vdots \\ w_s \end{bmatrix}}_{\mathbf{w}}$$

- Discriminative power of global histogram is low

# Weighted Histograms in Local Areas

- Accumulate into local areas to maintain different body parts




accumulation area

$$\mathbf{h}_1 = \sum_r \mathbf{f}_r w_r$$

$$\mathbf{h}_2 = \sum_r \mathbf{f}_r w_r$$

$$\vdots$$

$$\mathbf{h}_g = \sum_r \mathbf{f}_r w_r$$

local weighted histograms

$$\underbrace{\begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \vdots \\ \mathbf{h}_g \end{bmatrix}}_{\boldsymbol{x}} = \underbrace{\begin{bmatrix} \mathbf{f}_1 & \cdots & \mathbf{f}_a & 0 & \cdots & \cdots & 0 \\ * & \cdots & \cdots & \cdots & \cdots & \cdots & * \\ & & & \vdots & & & \\ 0 & \cdots & \cdots & 0 & \mathbf{f}_b & \cdots & \mathbf{f}_s \end{bmatrix}}_{\boldsymbol{F}} \underbrace{\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_s \end{bmatrix}}_{\boldsymbol{w}}$$

feature vector

feature matrix

weight map vector

Input to metric learning

# Mahalanobis Distance Metric Learning

- Squared Mahalanobis-like distance

$$D^2_M(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{M} (\boldsymbol{x}_i - \boldsymbol{x}_j)$$

  - Estimate $\boldsymbol{M}$ to rescale the feature vector
  - $\boldsymbol{M} = \boldsymbol{\Sigma}^{-1}$: Mahalanobis distance

Our case

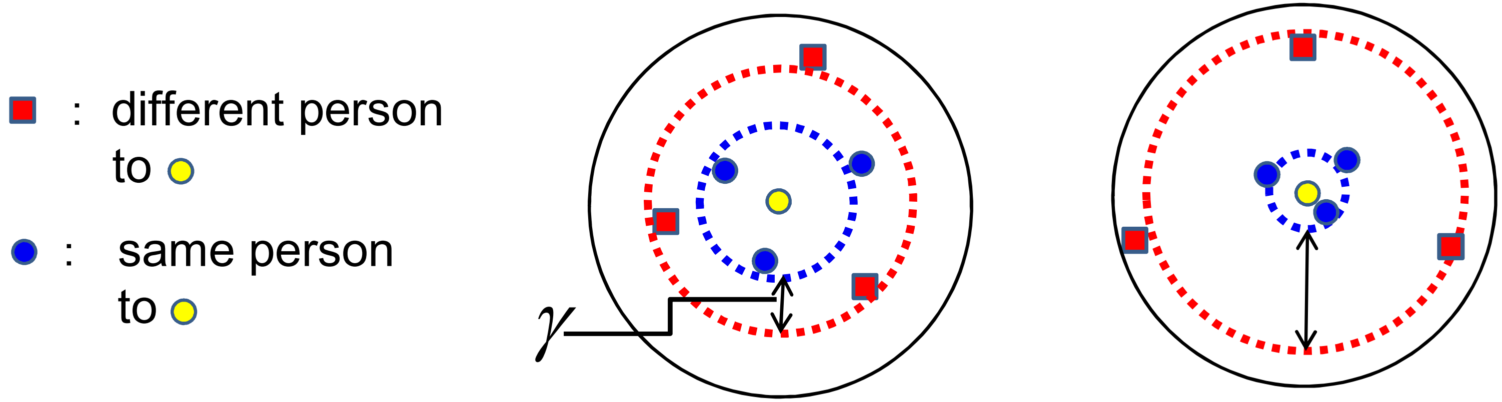
- Positive Semi-Definite Matrix $\boldsymbol{M} = \boldsymbol{L}\boldsymbol{L}^T$
- Definition of feature vector $\boldsymbol{x} = \boldsymbol{F}\boldsymbol{w}$

$$D^2_{w,L}(\boldsymbol{F}_i, \boldsymbol{F}_j) = \|\boldsymbol{L}^T \boldsymbol{F}_i \boldsymbol{w} - \boldsymbol{L}^T \boldsymbol{F}_j \boldsymbol{w}\|_2^2$$

- Estimate $\{\boldsymbol{w}, \boldsymbol{L}\}$ to transform the feature matrix

# Average Neighborhood Margin Maximization [Wang09]

- Explore discriminative information locally on sample space




original　　objective

Average Neighborhood Margin

$$\gamma_i = \underbrace{\sum_{j \in \mathcal{N}_i^D} \frac{D_{w,L}^2(\boldsymbol{F}_i, \boldsymbol{F}_j)}{|\mathcal{N}_i^D|}}_{\text{avg. neighborhood distances to different persons}} - \underbrace{\sum_{j \in \mathcal{N}_i^S} \frac{D_{w,L}^2(\boldsymbol{F}_i, \boldsymbol{F}_j)}{|\mathcal{N}_i^S|}}_{\text{avg. neighborhood distances to same person}}$$

# Optimization Problem

- Consider multiple pairs of weight map and metric $\{\boldsymbol{w}_k, \boldsymbol{L}_k\}_{k=1}^{K}$

- Objective $\sum_{k=1}^{K} J(\boldsymbol{w}_k, \boldsymbol{L}_k) = \sum_{i=1}^{N}\sum_{k=1}^{K} \gamma_{i,k}$ : summing up margins of all training samples and pairs

- Optimization problem becomes:

$$
\begin{aligned}
\max_{\boldsymbol{W},\boldsymbol{L}} \quad & \sum_{k=1}^{K} J(\boldsymbol{w}_k, \boldsymbol{L}_k) \\
s.t. \quad & \boldsymbol{W}^T\boldsymbol{W} = \boldsymbol{I}, \\
& \boldsymbol{L}_k^T\boldsymbol{L}_k = \boldsymbol{I},\ k = 1, \dots, K
\end{aligned}
$$

: un-correlate of weight map

: valid metric

  - Difficult to get the global solution
  - Solve it by a greedy algorithm for approximation

# Greedy Solution

- Separate into K steps and then sequentially solve them
  - Optimization problem of k-th step

$$\max_{\boldsymbol{w}_k, \boldsymbol{L}_k} \quad J(\boldsymbol{w}_k, \boldsymbol{L}_k)$$

: optimize k-th pair

$$s.t. \quad \boldsymbol{w}_k^T \boldsymbol{w}_k = 1,$$

$$\boxed{\boldsymbol{w}_k^T \boldsymbol{w}_m = 0,\ m = 1, ..., k-1,}$$

: un-correlate to learned weight maps

$$\boldsymbol{L}_k^T \boldsymbol{L}_k = \boldsymbol{I}.$$

  - Simplify by un-correlation of feature matrix [Appendix]

$$\boldsymbol{F}'_i \leftarrow \boldsymbol{F}_i - \sum_{m=1}^{k-1} \{(\boldsymbol{1}_d \otimes \boldsymbol{w}_m^T) \odot (\boldsymbol{1}_s^T \otimes \boldsymbol{F}_i \boldsymbol{w}_m)\}$$

$\otimes$ : kroneker product
$\odot$ : element-wise product

  - Simplified optimization problem of k-th step

$$\max_{\boldsymbol{w}_k, \boldsymbol{L}_k} \quad J'(\boldsymbol{w}_k, \boldsymbol{L}_k)$$

$$s.t. \quad \boldsymbol{w}_k^T \boldsymbol{w}_k = 1,$$

$$\boldsymbol{L}_k^T \boldsymbol{L}_k = \boldsymbol{I}.$$

➢ Solve it by alternative optimization

# Alternative Optimization for k-th Step A

- Repeat the steps A) and B) several times

A) Optimize $\boldsymbol{L}_k$ by fixing $\boldsymbol{w}_k$

- Transform training feature matrices: $\{\boldsymbol{x}_i = \boldsymbol{F}_i' \boldsymbol{w}_k\}_{i=1}^N$
- Search neighborhood sets for each sample
- Optimization problem becomes:

$$\boldsymbol{L}_k^* = \underset{\boldsymbol{L}_k}{\operatorname{argmax}}\, Tr\{\boldsymbol{L}_k^T (\boldsymbol{\Sigma}_D^{w_k} - \boldsymbol{\Sigma}_S^{w_k}) \boldsymbol{L}_k\}\ s.t.\ \boldsymbol{L}_k^T \boldsymbol{L}_k = \boldsymbol{I}$$

where
$$\boldsymbol{\Sigma}_D^{w_k} = \sum_{i=1}^{N} \sum_{j \in \mathcal{N}_{i,k}^D} \frac{(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T}{|\mathcal{N}_{i,k}^D|}$$

$$\boldsymbol{\Sigma}_S^{w_k} = \sum_{i=1}^{N} \sum_{j \in \mathcal{N}_{i,k}^S} \frac{(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T}{|\mathcal{N}_{i,k}^S|}$$

- Solve it by eigen value problem: $(\boldsymbol{\Sigma}_D^{w_k} - \boldsymbol{\Sigma}_S^{w_k}) \boldsymbol{L}_k = \lambda \boldsymbol{L}_k$

# Alternative Optimization for k-th Step B

- Repeat the steps A) and B) several times

B) Optimize $\boldsymbol{w}_k$ by fixing $\boldsymbol{L}_k$

- Transform training feature matrices: $\{\boldsymbol{Y}_i = \boldsymbol{L}_k^T \boldsymbol{F}'_i\}_{i=1}^N$
- Search neighborhood sets for each sample
- Optimization problem becomes:

$$\boldsymbol{w}_k^* = \underset{\boldsymbol{w}_k}{\operatorname{argmax}}\, \boldsymbol{w}_k^T (\boldsymbol{\Sigma}_D^{L_k} - \boldsymbol{\Sigma}_S^{L_k}) \boldsymbol{w}_k \quad s.t.\ \boldsymbol{w}_k^T \boldsymbol{w}_k = 1.$$

where $$\boldsymbol{\Sigma}_D^{L_k} = \sum_{i=1}^{N} \sum_{j \in \mathcal{N}_{i,k}^D} \frac{(\boldsymbol{Y}_i - \boldsymbol{Y}_j)^T (\boldsymbol{Y}_i - \boldsymbol{Y}_j)}{|\mathcal{N}_{i,k}^D|}$$

$$\boldsymbol{\Sigma}_S^{L_k} = \sum_{i=1}^{N} \sum_{j \in \mathcal{N}_{i,k}^S} \frac{(\boldsymbol{Y}_i - \boldsymbol{Y}_j)^T (\boldsymbol{Y}_i - \boldsymbol{Y}_j)}{|\mathcal{N}_{i,k}^S|}$$

- Solve it by eigen value problem: $(\boldsymbol{\Sigma}_D^{L_k} - \boldsymbol{\Sigma}_S^{L_k}) \boldsymbol{w}_k = \gamma \boldsymbol{w}_k$

# Experiments

- Three types of visual feature extracted from 15x15 grid cells
  - HSV color histogram:  24 dimension/cell
  - Gradient orientation histogram in YBbCr: 24 dimension/cell
  - Texture  (13 Schmid, 6 Gabor)  histogram: 152 dimension/cell

- Show generality with four public datasets

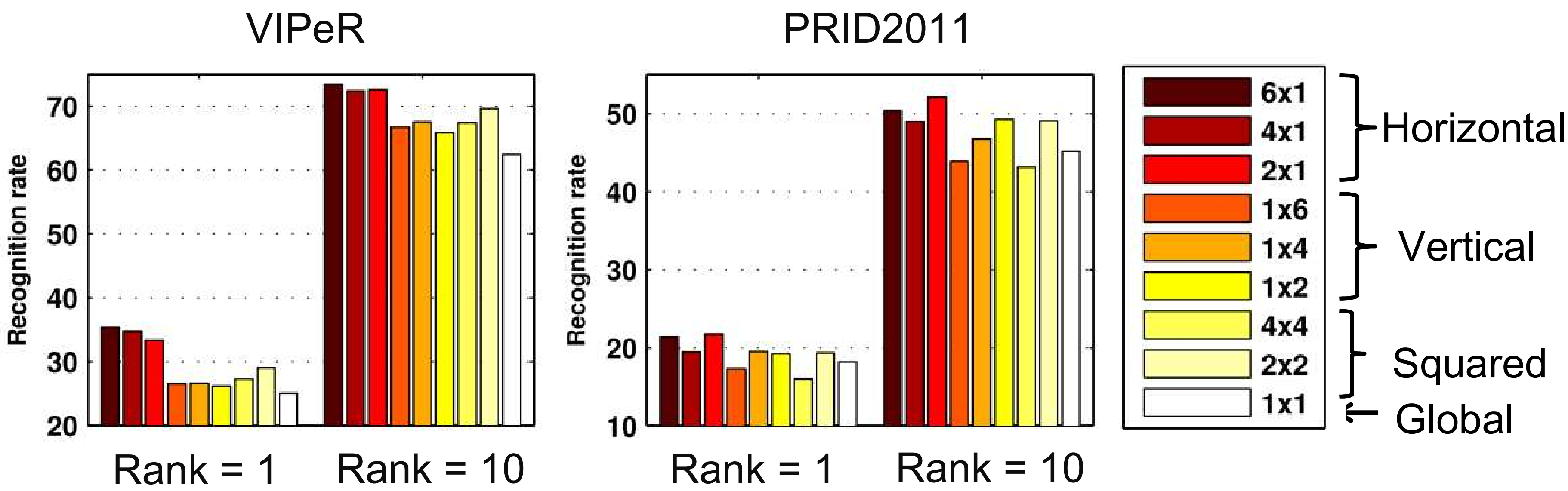
VIPeR

PRID2011

GRID

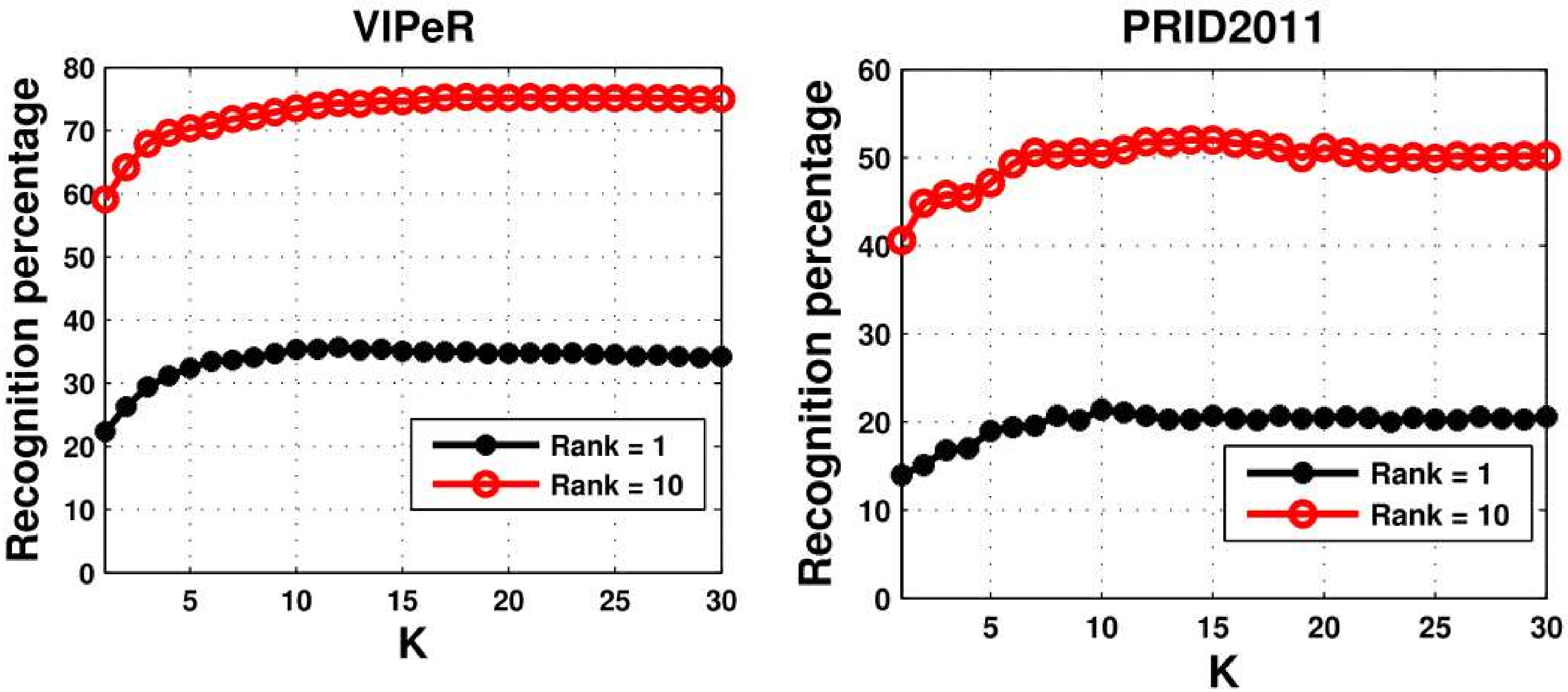
CAVIAR

# Comparison1 Accumulation Areas




- More areas than global histogram are better
- Horizontal strips are better than vertical strips
- Used 6x1 as default setting

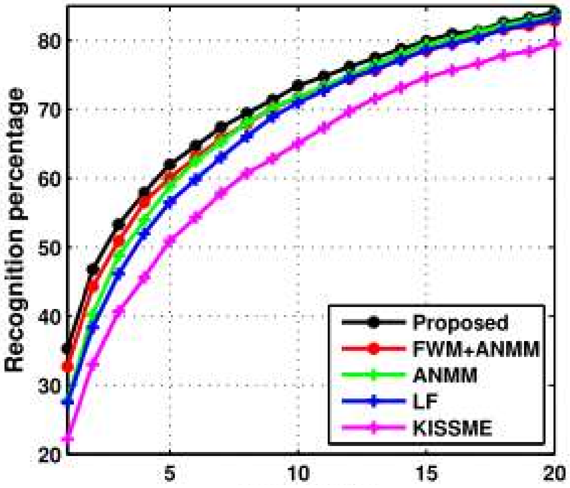
# Comp.2 Number of Weight Map/Metric Pairs



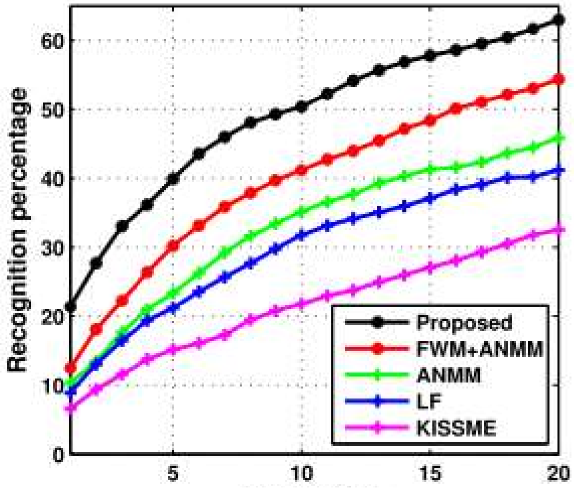

- Performance increases as increase the number of pairs
- Saturating around K=10
  - Used this setting as default

# Comp.3 Different Methods with Same Features

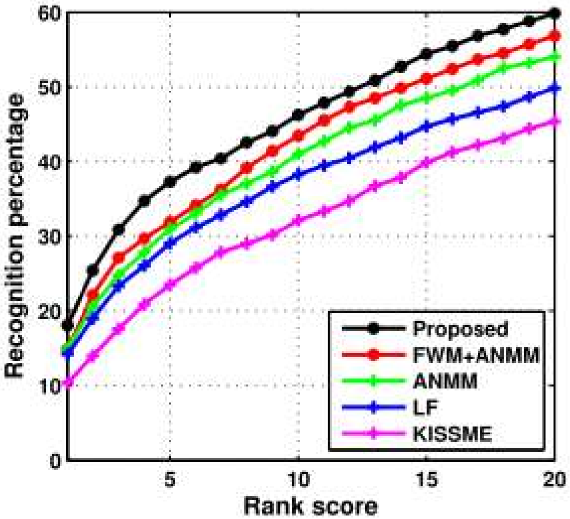
VIPeR



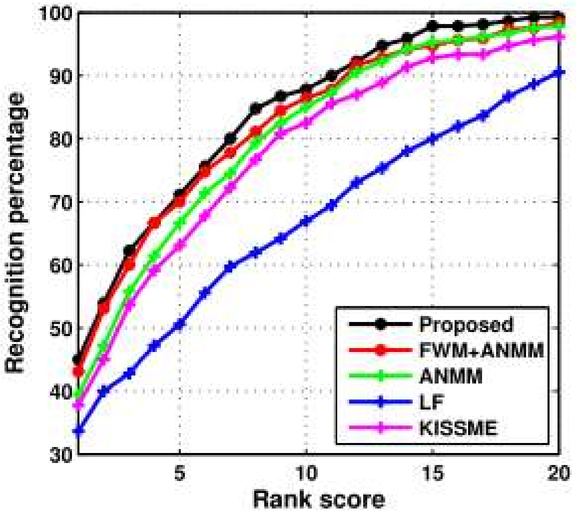

PRID2011




GRID




CAVIAR




- Compared with
  - *FWM* [Shinohara04]
  - *ANMM* [Wang09]
  - *KISSME* [Kostinger12]
  - *LF* [Pedagadi13]

# Comparison 4 Other Reported Results

| Rank score | 1 | 5 | 10 | 20 |
|---|---|---|---|---|
| VIPeR | | | | |
| Proposed | **35.35** | **62.03** | **73.48** | **84.05** |
| *RPLM* [Hirzer12] | 27 | 60 | 69 | 83 |
| *RDC* [Zheng12] | 15.6 | 38.42 | 53.8 | 70.09 |
| *SCEFA* [Hu13] | 26.49 | 49.80 | 60.29 | 73.54 |
| *SDALF* [Bazzani13] | 19.11 | 38.97 | 51.07 | 65.29 |
| PRID2011 | | | | |
| Proposed | **21.4** | **39.9** | **50.4** | **63.0** |
| *RPLM* [Hirzer12] | 15 | 33 | 42 | 54 |
| GRID | | | | |
| Proposed | **18.08** | **37.28** | **46.24** | **59.84** |
| *MRank* [Loy13] | 12.24 | 27.84 | 36.32 | 46.56 |
| CAVIAR | | | | |
| Proposed | **45.00** | **71.11** | **87.78** | **99.17** |
| *LF* [Pedagadi13] | 33.61 | 50.55 | 66.94 | 90.55 |

# Comparison 5 Small Number of Training Person

VIPeR

| Rank score | 1 | 5 | 10 | 20 |
|---|---|---|---|---|
| # of training persons = 200 (gallery size 432) | | | | |
| Proposed | **25.93** | **50.32** | **63.19** | **76.3** |
| *RDC* [Zheng12] | 12.29 | 31.55 | 44.49 | 59.91 |
| *SCEFA* [Hu13] | 23.71 | 45.39 | 55.39 | 67.89 |
| *SDALF* [Bazzani13] | 16.58 | 34.8 | 45.09 | 58.75 |
| # of training persons = 100 (gallery size 532) | | | | |
| Proposed | 20.0 | 40.92 | **53.46** | **66.67** |
| *RPLM* [Hirzer12] | 11 | 25 | 38 | 52 |
| *RDC* [Zheng12] | 9.12 | 24.19 | 34.40 | 48.55 |
| *SCEFA* [Hu13] | **22.13** | **42.72** | 52.3 | 63.19 |
| *SDALF* [Bazzani] | 15.19 | 31.72 | 41.45 | 54.15 |

metric learning

sophisticated features

- Metric learning was weak when small size of training persons
- Proposed method is comparable to sophisticated features

# Results of Weight Analysis

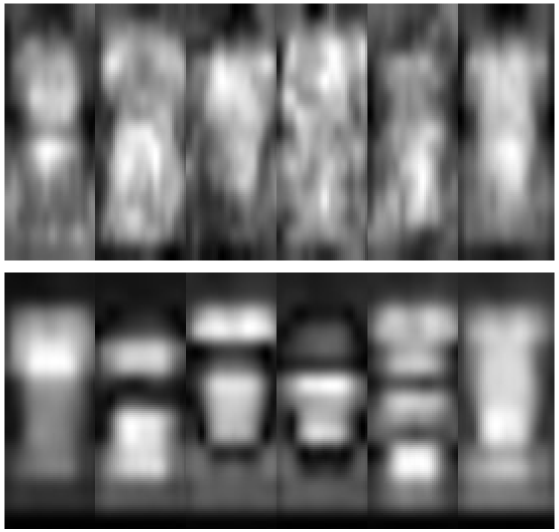
VIPeR

PRID2011

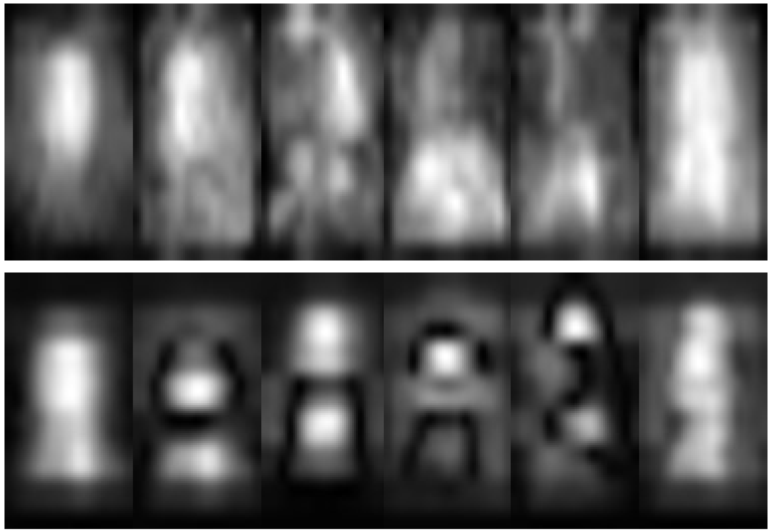
Dataset

Conventional

Proposed



1 2 3 4 5 avg



1 2 3 4 5 avg

# Conclusion

- Discriminative Accumulation of Local Features for Person Re-Identification
  - Jointly learn multiple pairs of metric and weight map
  - Superior performance to state-of-the arts methods

- Future work
  - Extension to be adaptive for each input image