Person Identification: from Face to Body

Wei-Shi Zheng, Jian-Huang Lai

SUN YAT-SEN UNIVERSITY
Start with A Video
The Two Problems

Body

Face
The Two Problems

Person Re-identification

Face Recognition
Outline of This Talk

- Our work on Face Recognition
  - Discriminant Feature Extraction: A Perturbation Model
  - Pre-image Problem in KPCA
- Our work on Person Re-identification
  - Pedestrian Detection with Context
  - A Distance Comparison based Person Re-identification model
Face Recognition: Our focus

- Illumination
- Pose Variation
- Occlusion/Corruption
- Expression variation
Our work on Face Recognition

• Pose Invariant Feature Extraction
  – Lai and Yuen, Pattern Recognition 2001
• Discriminant Feature Extraction by Subspace Analysis
• Independent Sparse Feature Extraction
  – Yuen and Lai, Pattern Recognition 2003
  – Zheng et al., ICCV 2007
• Kernel Methods for Nonlinear Facial Image Processing
  – Zheng, Lai et al., IEEE TNN 2010
• Illumination Normalization
Our work on Face Recognition

• Pose Invariant Feature Extraction
  – Lai and Yuen, Pattern Recognition 2001

• Discriminant Feature Extraction by Subspace Analysis

• Independent / Sparse Feature Extraction
  – Yuen and Lai, Pattern Recognition 2003
  – Zheng et al., ICCV 2007

• Kernel Methods for Nonlinear Facial Image Processing
  – Zheng, Lai et al., IEEE TNN 2010

• Illumination Normalization
Discriminant Feature Extraction

- Our research focus: Fisher Criterion

\[ W_{opt} = \arg \max_W \frac{\text{trace}(W^T S_b W)}{\text{trace}(W^T S_w W)}. \]

\[ S_w^{-1} S_b W = W \Lambda \]

- Between-class covariance matrix
- Within-class covariance matrix
• **Our concern: Small Sample Size Problem**
  – Limited samples for each person
  – The within-class scatter matrix $S_w$ can be singular, and its inverse does not exist

$$W_{opt} = \arg \max_{W} \frac{\text{trace}(W^T S_b W)}{\text{trace}(W^T S_w W)}.$$  \[ \iff \]  $$S_w^{-1} S_b W = W \Lambda$$

• **Our investigations:**
  – GA-Fisher: Selecting Principal Components
  – 2D-LDA vs. 1D-LDA
  – Perturbation Models for 1D-LDA
Discriminant Feature Extraction

• Early work I: GA-Fisher (Zheng & Lai, IEEE TSMC-B)
  – Selecting the most discriminant and illumination invariant principal component vectors in LDA+PCA by a proposed Genetic Algorithm.

• LDA+PCA:

\[ W_{opt} = W_{fld} W_{pca} \]

\[ W_{fld} = \arg \max_W \frac{|W^T \hat{S}_b W|}{|W^T \hat{S}_w W|} \]
Discriminant Feature Extraction

• **Early work II: 1D-LDA vs. 2D-LDA** (Zheng & Lai, PR 2008)
  – 1D-LDA: Using Image Vectors as Input
  – 2D-LDA: Using Image Matrices as Input

  **Using Geometric Information**
  \[
  \mathbf{w}_l \times \begin{bmatrix}
  1 & 4 & 7 \\
  2 & 5 & 8 \\
  3 & 6 & 9 \\
\end{bmatrix} \times \begin{bmatrix}
  1 & 4 & 7 \\
  2 & 5 & 8 \\
  3 & 6 & 9 \\
\end{bmatrix} = \begin{bmatrix}
  1 & 2 & 3 \\
  4 & 5 & 6 \\
  7 & 8 & 9 \\
\end{bmatrix}
  \]

  **Two-dimensional Approach**

  **Traditional Approach**

  - 2D-LDA actually loses the cross-covariance information between rows or columns
  - 2D-LDA is sensitive to the number of features
  - Regularized 1D-LDA could always perform better
Perturbation LDA (Zheng & Lai, PR 2009)

- Regularized LDA always performs very well

\[ W_{opt} = \arg\max_w \frac{\text{trace}(W^T \hat{S}_b W)}{\text{trace}(W^T (\hat{S}_w + \lambda I) W)}, \quad \lambda > 0 \]

\[
\hat{S}_w = \sum_{k=1}^L \frac{N_k}{N} \hat{S}_k, \quad \hat{S}_k = \sum_{i=1}^{N_k} \frac{1}{N_k} (\mathbf{x}_i^k - \hat{\mathbf{u}}_k)(\mathbf{x}_i^k - \hat{\mathbf{u}}_k)^T
\]

\[
\hat{S}_b = \frac{1}{2} \sum_{k=1}^L \sum_{j=1}^L \frac{N_k}{N} \times \frac{N_j}{N} (\hat{\mathbf{u}}_k - \hat{\mathbf{u}}_j)(\hat{\mathbf{u}}_k - \hat{\mathbf{u}}_j)^T
\]

- We find the regularization can describe the perturbation between an empirical class center and its expectation

- We develop a complete perturbation model for fisher criterion

\[ \tilde{W}_{opt} = \arg\max_w \frac{\text{trace}(W^T (\hat{S}_b + \hat{S}_b^\lambda) W)}{\text{trace}(W^T (\hat{S}_w + \hat{S}_w^\lambda) W)} \]
Motivation

Fisher’s LDA makes different class centers far away from each other and data of the same class close to their corresponding class centers.

\[ S_w = \frac{1}{N} \sum_{k=1}^{L} \sum_{i=1}^{N_k} (x_i^k - u_k)(x_i^k - u_k)^T \]
\[ S_b = \frac{1}{2} \sum_{k=1}^{L} \sum_{j=1}^{N_k} \frac{N_k}{N} \times \frac{N_j}{N} (u_k - u_j)(u_k - u_j)^T \]

The estimate of the class mean cannot be accurate in the undersampled case.

Inaccurate estimation of the class mean would make LDA degraded.
Perturbation Model

A random vector is first formulated for any sample \( x \) to stochastically approximate \( E_{x'|y}[x'] \) below:

\[
\tilde{x} = x + \xi_x, \quad \xi_x \sim \mathcal{N}(0, \Omega_y), \quad \Omega_y \in \mathbb{R}^{n \times n}
\]

The stochastic approximation of \( \tilde{x} \) to \( E_{x'|y}[x'] \) means there exists a specified estimate \( \hat{\xi}_x \) of the random vector \( \xi_x \) with respect to the corresponding distribution such that

\[
x + \hat{\xi}_x = E_{x'|y}[x']
\]

Model the random mean of each class:

\[
\hat{u}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} \tilde{x}_i^k = u_k + \frac{1}{N_k} \sum_{i=1}^{N_k} \xi_i^k
\]
New Estimations of Covariance Matrices

\[ \tilde{S}_b = E_\xi \left[ \frac{1}{2} \sum_{k=1}^L \sum_{j=1}^L \frac{N_k}{N} \times \frac{N_j}{N} (\tilde{u}_k - \tilde{u}_j)(\tilde{u}_k - \tilde{u}_j)^T \right] = S_b + S^A_b, \]

result in some proper correctings (perturbations) on the empirical between-class and within-class covariance matrices.

\[ \tilde{S}_w = \sum_{k=1}^L \frac{N_k}{N} E_\xi \left[ \sum_{i=1}^{N_k} \frac{1}{N_k} (x_i^k - \tilde{u}_k)(x_i^k - \tilde{u}_k)^T \right] = S_w + S^A_w, \]
**Perturbation LDA**

- Develop perturbation model for learning the impact of the difference between an empirical class mean and its expectation in Fisher criterion.

- Develop the Perturbation LDA

  \[
  \tilde{W}_{opt} = \arg \max_w \frac{\text{trace}(W^T (\hat{S}_b + \hat{S}_b^\Delta) W)}{\text{trace}(W^T (\hat{S}_w + \hat{S}_w^\Delta) W)}
  \]

- Develop a regularized estimation method for estimating the perturbation in the whole PCA space.
Develop the Semi-Perturbation LDA

\[ S_{b}^{\Lambda} \text{ is ignored.} \]

A novel view to Regularized LDA

\[
\tilde{W}_{opt} = \arg \max_w \frac{\text{trace}(W^T \hat{S}_b W)}{\text{trace}(W^T (\hat{S}_w + S_{w}^{\Lambda}) W)}
\]

Efficient parameter estimation for Regularized LDA
Some Experiments

### Perturbation LDA

<table>
<thead>
<tr>
<th>Method</th>
<th>Classifier: NCMC</th>
<th>Classifier: NNC</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-LDA</td>
<td>87.06%</td>
<td>89.29%</td>
</tr>
<tr>
<td>R-LDA (CV) [37]</td>
<td>86.43%</td>
<td>87.96%</td>
</tr>
<tr>
<td>N-LDA [12]</td>
<td>83.49%</td>
<td>83.49%</td>
</tr>
<tr>
<td>Direct LDA [34]</td>
<td>80.71%</td>
<td>78.98%</td>
</tr>
<tr>
<td>Fisherface [1]</td>
<td>77.25%</td>
<td>71.22%</td>
</tr>
</tbody>
</table>

#### Average Recognition Accuracy on Subset of FERET (p=3)

<table>
<thead>
<tr>
<th>Method</th>
<th>Classifier: NCMC</th>
<th>Classifier: NNC</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-LDA</td>
<td>81.82%</td>
<td>93.26%</td>
</tr>
<tr>
<td>R-LDA (CV) [37]</td>
<td>80.43%</td>
<td>93.29%</td>
</tr>
<tr>
<td>N-LDA [12]</td>
<td>74.45%</td>
<td>84.98%</td>
</tr>
<tr>
<td>Direct LDA [34]</td>
<td>72.73%</td>
<td>88.12%</td>
</tr>
<tr>
<td>Fisherface [1]</td>
<td>67.26%</td>
<td>82.17%</td>
</tr>
</tbody>
</table>

#### Average Recognition Accuracy on Subset of CMU PIE

<table>
<thead>
<tr>
<th>Method</th>
<th>Classifier: NCMC</th>
<th>Classifier: NNC</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-LDA</td>
<td>78.98%</td>
<td>93.26%</td>
</tr>
<tr>
<td>R-LDA (CV) [37]</td>
<td>78.44%</td>
<td>93.29%</td>
</tr>
<tr>
<td>N-LDA [12]</td>
<td>74.45%</td>
<td>84.98%</td>
</tr>
<tr>
<td>Direct LDA [34]</td>
<td>73.68%</td>
<td>88.12%</td>
</tr>
<tr>
<td>Fisherface [1]</td>
<td>72.99%</td>
<td>82.17%</td>
</tr>
</tbody>
</table>
## Some Experiments

### Average Recognition Accuracy of R-LDA on FERET Data Set: “R-LDA with manually selected optimal parameter” vs. “R-LDA using perturbation model” (p=3)

<table>
<thead>
<tr>
<th>Method</th>
<th>Classifier: NCMC</th>
<th></th>
<th></th>
<th>Classifier: NNC</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rank 1</td>
<td>Rank 2</td>
<td>Rank 3</td>
<td>Rank 1</td>
<td>Rank 2</td>
<td>Rank 3</td>
</tr>
<tr>
<td>R-LDA with manually selected optimal parameter</td>
<td>86.78%</td>
<td>90.24%</td>
<td>91.69%</td>
<td>88.27%</td>
<td>90.16%</td>
<td>91.25%</td>
</tr>
<tr>
<td>R-LDA (CV)</td>
<td>86.43%</td>
<td>89.96%</td>
<td>91.49%</td>
<td>87.96%</td>
<td>90.26%</td>
<td>91.33%</td>
</tr>
<tr>
<td>R-LDA using perturbation model</td>
<td>86.47%</td>
<td>90.00%</td>
<td>91.69%</td>
<td>88.08%</td>
<td>90.20%</td>
<td>91.49%</td>
</tr>
</tbody>
</table>

### Average Recognition Accuracy of R-LDA on CMU PIE Data Set: “R-LDA with manually selected optimal parameter” vs. “R-LDA using perturbation model” (p=5)

<table>
<thead>
<tr>
<th>Method</th>
<th>Classifier: NCMC</th>
<th></th>
<th></th>
<th>Classifier: NNC</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rank 1</td>
<td>Rank 2</td>
<td>Rank 3</td>
<td>Rank 1</td>
<td>Rank 2</td>
<td>Rank 3</td>
</tr>
<tr>
<td>R-LDA with manually selected optimal parameter</td>
<td>78.60%</td>
<td>83.42%</td>
<td>85.88%</td>
<td>80.50%</td>
<td>84.08%</td>
<td>85.98%</td>
</tr>
<tr>
<td>R-LDA (CV)</td>
<td>78.44%</td>
<td>83.27%</td>
<td>85.72%</td>
<td>80.43%</td>
<td>84.05%</td>
<td>85.94%</td>
</tr>
<tr>
<td>R-LDA using perturbation model</td>
<td>78.24%</td>
<td>83.51%</td>
<td>86.13%</td>
<td>80.18%</td>
<td>84.12%</td>
<td>86.14%</td>
</tr>
</tbody>
</table>
Pre-image Learning in KPCA

- Penalized Pre-image Learning in KPCA
The Pre-image Problem in KPCA

The goal is to find a data point \( \hat{x} \) such that \( \phi(\hat{x}) = P_k \phi(x) \).

The problem: \( \phi(\hat{x}) \neq P_k \phi(x) \) always.

The exact pre-image does not always exist. It is an ill-posed problem and currently an open issue!!!
The Proposed Two-step Framework

\[ \hat{x} = \Psi w_x = \sum_{i=1}^{N} w_i^x x_i, \]

where

\[ \begin{align*}
  w_x &= \arg \min \ G(w \mid P_x \varphi(x), \{x_i\}), \\
  w &= (w_1, \ldots, w_N)^T \in \mathbb{R}^N \\
  s.t. \quad w_x & \text{ satisfies some constraints}
\end{align*} \]

Range of Pre-image: Combination of training samples

Our approach: Explicitly

\[ \text{VS.} \]

Previous: Implicitly
P2L (Penalized Pre-image Learning)

\[
\hat{x} = \Psi_{s,x} w_x = \sum_{i=1}^{s} w_i^x \hat{x}_i,
\]

其中

\[
\begin{align*}
    w_x &= \arg\min_{w=(w_1, \ldots, w_s)^T \in \mathbb{R}^s} w^T \Psi_{s,\varphi(x)}^T \Psi_{s,\varphi(x)} w - 2(\Psi_{x,\gamma}^x)^T \Psi_{s,\varphi(x)} w + \lambda \cdot F(w) \\
    s.t. &\quad \sum_{i=1}^{s} w_i^x = 1 \quad \& \quad w_i^x \geq 0, \quad i = 1, \ldots, s, \quad \lambda \geq 0
\end{align*}
\]

Integration of Prior Knowledge

Penalization

make pre-image well-defined
Pre-image Learning in KPCA

- A weakly supervised penalty for guiding the learning process

\[
F(w) = \eta^+ |H^+_{\theta}(x)|^{-1} \left[ \sum_{z^+ \in H^+_{\theta}(x)} \|\Psi_{s,x} w - z^+\|^2 \right] \\
- \eta^- |H^-_{\theta}(x)|^{-1} \left[ \sum_{z^- \in H^-_{\theta}(x)} \|\Psi_{s,x} w - z^-\|^2 \right]
\]
Pre-image Learning in KPCA

• Expression Normalization

(a) Facial Expression Images

(b) References of Natural Facial Expression Images

(e) Distance Constraint

(g) Penalized Pre-image Learning (Weakly Supervised)
Pre-image Learning in KPCA

- Occlusion Recovery
Illumination Problem

- Small Scale and Large Scale Feature Normalization
  (Xie, Zheng, Lai et al., IEEE TIP 2011)
Illumination Problem

illuminated images

small scale

large scale

normalization on the whole image by NPL-QI

normalization using proposed framework without compensation

normalization using proposed framework with compensation
The Two Problems

Person Re-identification

Face Recognition
Basic Scenario for Visual Surveillance

Detecting target objects (Cars, pedestrian, bags, etc)

Matching, Tracking

What is he doing?

Camera Network Understanding
Our Current Focus in Visual Surveillance

- Detecting target objects (Cars, pedestrian, bags)
- Matching, Tracking
- Camera Network Understanding

What is he doing?
Our Work

• Pedestrian Detection with Context
  – Zheng et al. ICCV 2009, TPAMI 2011 (accepted)

• Person Re-identification, Consistent Labelling
  – Zheng et al. CVPR 2011
  – Lian, Lai, Zheng, Pattern Recognition 2011

• Group Association
  – Zheng, BMVC 2009
Pedestrian Detection with Context

• Describing the contextual information:
  – A Polar Geometric Context Descriptor
Pedestrian Detection with Context

- Quantifying context
  - Not all context is useful and discriminant
  - Using the selective context to improve the prior detection confidence provided by any basic object detector

\[ \sum \delta(s_i^\alpha \cdot g(h_i) \leq s_j^\alpha \cdot g(h_j)) \]
Pedestrian Detection with Context

Prior Detection Using HOG detector

With MMC model
Pedestrian Detection with Context

• Transferring Context
  – Context is diverse
  – May not have enough samples for quantifying context for each target class.
  – Different classes may share similar surrounding information
  – Or the general effect of context helps improve object detection for different classes is similar
A. TMMC-I: Transferring Discriminant Contextual Information

\[
\begin{align*}
    w^T \varphi(h_i) + b_q + \alpha_q \cdot \log s_i & \geq \rho_q - \xi_i, \; \forall \; y_i = 1 \text{ and } \tau_i = q, \\
    w^T \varphi(h_j) + b_q + \alpha_q \cdot \log s_j & \leq -\rho_q + \xi_j, \; \forall \; y_j = -1 \text{ and } \tau_j = q.
\end{align*}
\]

\[
\{w_t, b^t_q, \alpha^t_q\} = \arg \min_{w, b_q, \alpha_q, \rho_q, \xi_i} \frac{1}{2} \left( \|w\|^2 + \sum_{q=1}^{Q} \alpha^2_q \right) + \frac{1}{N} \sum_{i=1}^{N} \xi_i - \frac{1}{N} \sum_{q=1}^{Q} N_q \cdot \rho_q 
\]

\[
s.t. \quad y_i (w^T \varphi(h_i) + b_q + \alpha_q \log s_i) \geq \rho_q - \xi_i, \; if \; \tau_i = q,
\]

\[
\xi_i, \; \rho_q \geq 0.
\]
Pedestrian Detection with Context

• Transfer Maximum Margin Context (TMMC) Model

B. TMMC-II: Transferring the Weight of Prior Detection Score

\[
\begin{align*}
\mathbf{w}_q^T \varphi(h_i) + b_q + \alpha \cdot \log s_i & \geq \rho_q - \xi_i, \quad \forall \ y_i = 1 \text{ and } \tau_i = q, \\
\mathbf{w}_q^T \varphi(h_j) + b_q + \alpha \cdot \log s_j & \leq -\rho_q + \xi_j, \quad \forall \ y_j = -1 \text{ and } \tau_j = q.
\end{align*}
\]

\[
\{ \mathbf{w}_q^t, b_q^t, \alpha_t \} = \arg \min_{\mathbf{w}_q, b_q, \alpha, \rho_q, \xi_i} \frac{1}{2} \left( \sum_{q=1}^{Q} \| \mathbf{w}_q \|^2 + \alpha^2 \right) + \frac{1}{N} \sum_{i=1}^{N} \xi_i - \frac{\nu}{N} \sum_{q=1}^{Q} N_q \cdot \rho_q
\]

\[
\text{s.t. } y_i (\mathbf{w}_q^T \varphi(h_i) + b_q + \alpha \log s_i) \geq \rho_q - \xi_i, \quad \text{if } \tau_i = q,
\]

\[
\xi_i, \rho_q \geq 0.
\]
### Effect of transfer learning

<table>
<thead>
<tr>
<th>Target Category</th>
<th>Auxiliary Category</th>
<th>Model</th>
<th>Average Precision Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>car</td>
<td>motorbike</td>
<td>TMMC-I</td>
<td>0.4006 (0.3741)</td>
</tr>
<tr>
<td>motorbike</td>
<td>car</td>
<td>TMMC-I</td>
<td>0.4253 (0.4020)</td>
</tr>
<tr>
<td>vehicle</td>
<td>car</td>
<td>TMMC-I</td>
<td>0.3952 (0.3838)</td>
</tr>
<tr>
<td>people</td>
<td>car</td>
<td>TMMC-I</td>
<td>0.3963 (0.3862)</td>
</tr>
<tr>
<td>people</td>
<td>bicycle</td>
<td>TMMC-I</td>
<td>0.3887 (0.3862)</td>
</tr>
<tr>
<td>people</td>
<td>motorbike</td>
<td>TMMC-I</td>
<td>0.3916 (0.3862)</td>
</tr>
<tr>
<td>bicycle</td>
<td>people</td>
<td>TMMC-I</td>
<td>0.2703 (0.2878)</td>
</tr>
<tr>
<td>car</td>
<td>people</td>
<td>TMMC-I</td>
<td>0.3667 (0.3741)</td>
</tr>
<tr>
<td>motorbike</td>
<td>people</td>
<td>TMMC-I</td>
<td>0.3954 (0.4020)</td>
</tr>
<tr>
<td>bicycle</td>
<td>people</td>
<td>TMMC-II</td>
<td>0.3063 (0.2878)</td>
</tr>
<tr>
<td>people</td>
<td>bicycle</td>
<td>TMMC-II</td>
<td>0.3964 (0.3862)</td>
</tr>
<tr>
<td>vehicle</td>
<td>car</td>
<td>TMMC-II</td>
<td>0.4019 (0.3838)</td>
</tr>
<tr>
<td>car</td>
<td>motorbike</td>
<td>TMMC-II</td>
<td>0.3724 (0.3741)</td>
</tr>
<tr>
<td>motorbike</td>
<td>car</td>
<td>TMMC-II</td>
<td>0.3835 (0.4020)</td>
</tr>
<tr>
<td>car</td>
<td>people</td>
<td>TMMC-II</td>
<td>0.3720 (0.3741)</td>
</tr>
<tr>
<td>motorbike</td>
<td>people</td>
<td>TMMC-II</td>
<td>0.3932 (0.4020)</td>
</tr>
</tbody>
</table>
Person Re-identification

- Cross-camera Views (Non-overlapping)
Person Re-identification

• Large Intra-class/Inter-class Variations (limited samples)
Person Re-identification

- Existing methods for Person re-identification
  - Manually define robust representation
    - e.g. Spatial Graph, Appearance Context Model
  - Non-learning based distance
    - L1-Norm, Bhattacharyya distance
  - Adaboost for selecting discriminant features
    - Maximize the between-class distance
    - Minimize the within-class distance

- Problems:
  - Appearance of people:
    - Large intra-class variations + Large inter-class variations
  - Limited Samples
  - Traditional Discriminant Analysis (e.g. Adaboost) would incur overfitting
Person Re-identification

- Our solution: Proposing a soft discriminant distance learning model (Zheng et al. CVPR 2011):

\[ f(x) = x^T M x, \quad M \succeq 0 \]

\[ P(f(x^p_i) < f(x^n_i)) = \left(1 + \exp \left\{ f(x^p_i) - f(x^n_i) \right\} \right)^{-1} \]

\[ f = \arg \min_f r(f, \emptyset), \]

\[ r(f, \emptyset) = -\log \left( \prod_{i} P(f(x^p_i) < f(x^n_i)) \right) \]
Person Re-identification

The proposed method
Consistent Labelling Across Camera View

• A Spatial Temporal Model (Lian, Lai, Zheng, PR 2011)
  – We further incorporate the time delay information
Consistent Labelling Across Camera View

- **A Spatial Temporal Model**
  - We further develop a Optimal Graph Matching based method for solving the group split/merge problem
Group Matching

- Associating Groups of People (Zheng, BMVC2009)
  - Associating Group of People vs. Individuals
  - Challenges

(a) Ambiguities from person re-identification in isolation

(b) Associating groups of people may reduce ambiguities in matching

(c) Difficult examples of associating groups of people
Group Matching

• Modelling
  – Main Contribution I: Group Descriptor
    • A rectangle ring descriptor: rotation invariant

\[
\text{intra ratio-occurrence map } H_i \\
H_i(a, b) = \frac{h_i(a)}{h_i(a) + h_i(b) + \varepsilon}
\]

\[
\text{inter ratio-occurrence maps } S_i \text{ and } G_i \\
G_i(a, b) = \frac{g_i(a)}{g_i(a) + h_i(b) + \varepsilon}, \quad S_i(a, b) = \frac{s_i(a)}{s_i(a) + h_i(b) + \varepsilon}, \quad g_i = \sum_{j=1}^{i-1} h_j, \quad s_i = \sum_{j=i+1}^{l} h_j
\]
Group Matching

- Modelling
  - Main Contribution I: Group Descriptor
    - A block based occurrence descriptor: for large non-center-rotational changes in people’s positions

\[
T^i_b = \{H^i_j\}^{4\gamma+1}_{j=0} \cup \{O^i_j\}^{2}_{j=1}
\]

Intra ratio-occurrence map \( H^i_j \) between visual words in each block region \( SB^i_j \)

\[
H^i(a, b) = \frac{h^i(a)}{h^i(a) + h^i(b) + \varepsilon}
\]

Inter ratio-occurrence maps \( O^i_j \) between block \( B_i \) and its complementary region \( SB^i_{4\gamma+1} \)

\[
O^i_1(a, b) = \frac{t^i(a)}{t^i(a) + z^i(b) + \varepsilon} \quad O^i_2(a, b) = \frac{z^i(a)}{z^i(a) + t^i(b) + \varepsilon}
\]

where \( z^i \) and \( t^i \) are the histograms of visual words of block \( B_i \) and image region \( SB^i_{4\gamma+1} \)
Group Matching

• Modelling

– Main Contribution II: Metric

\[
d(I_1, I_2) = d_r \left( \{ T^i_r(I_1) \}_{i=1}^{\ell}, \{ T^i_r(I_2) \}_{i=1}^{\ell} \right) + \alpha \cdot d_b \left( \{ T^i_b(I_1) \}_{i=1}^{m_1}, \{ T^i_b(I_2) \}_{i=1}^{m_2} \right), \quad \alpha \geq 0
\]

L_1 norm metric

(a) CRRRO Descriptor

CRRRO

BRO

top k-match metric

(b) BRO Descriptor

\[
d_b \left( \{ T^i_b(I_1) \}_{i=1}^{m_1}, \{ T^i_b(I_2) \}_{i=1}^{m_2} \right) = \min_{C,D} \left\{ k^{-1} \cdot ||AC - BD||_1 \right\},
\]

A \in \mathbb{R}^{q \times m_1}, B \in \mathbb{R}^{q \times m_2}, C \in \mathbb{R}^{m_1 \times k}, D \in \mathbb{R}^{m_2 \times k}

the \( i^{th} (i''^{th}) \) column of matrix A (B) is the vector representation of \( T^i_b(I_1) (T^i_b(I_2)) \)
each column \( c_j (d_j) \) of C (D) is an indicator vector and the columns of C (D) are orthogonal
Group Matching

CMC Curve

- Holistic Color Histogram
- Holistic Visual Word Histogram
- Concatenated Histogram (RGB)
- Concatenated Histogram (SIFT)
- CRRRO-BRO

Probe Image

Rank 1

Rank 2

Rank 3

Rank 4

Rank 5
Our Current Research Interests

• Face Recognition in Uncontrolled Environment

• Recognition/Categorization based on Transfer Learning (for Face, object images)

• Visual Surveillance: Matching problem (e.g. pedestrian) and event analysis
• Thank you!