Neighborhood Repulsed Metric Learning for Kinship Verification

Jiwen Lu
Advanced Digital Sciences Center, Singapore
Paper

Face Analysis Tasks

Face identification (access control/ surveillance)

Face verification (access control/surveillance)
Face Analysis Tasks

Facial expression recognition *(human-computer interaction)*

Facial age estimation *(visual advertisement/social media)*
Face Analysis Tasks

Head pose estimation *(human computer interaction)*

Gender classification *(social media analysis)*
Face Analysis Tasks

Facial beauty prediction (multimedia analysis)
Face Analysis Tasks

Kinship verification *(social media analysis)*

Father-Son (F-S)  
Father-Daughter (F-D)

Mother-Son (M-S)  
Mother-Daughter (M-D)
Related Work

Local Features + SVM (Fang2010 [1])

- 150 image pairs
- 50%: Caucasians
- 40%: Asians
- 7% African Americans
- 3% others;
- 40%: F-S
- 22%: F-D
- 13%: M-S
- 26%: M-D

Related Work

Transfer Learning (Xia2012 [2])

- 90 groups
- 3 images in each group

Related Work

Salient Feature + SVM (Guo2012 [3])

Datasets

- KinFaceW-I: 500 kinship image face pairs
Datasets

• KinFaceW-II:1000 kinship image face pairs
Datasets

- Statistics

- Publicly available: www.kinfacew.com
Datasets

Home

Welcome to Kinship Face in the Wild (KinFaceW), a database of face images collected for studying the problem of kinship verification from unconstrained face images. There are many potential applications for kinship verification such as family album organization, genealogical research, missing family members search, and social media analysis.

The aim of kinship verification is to determine whether there is a kin relation between a pair of given face images. The kinship is defined as a relationship between two persons who are biologically related with overlapping genes. Hence, there are four representative types of kin relations: Father-Son (F-S), Father-Daughter (F-D), Mother-Son (M-S) and Mother-Daughter (M-D), respectively.

News!

Sep-22-2014: The detailed information of The Kinship Verification in the Wild Evaluation can be found here, which is organized as part of FG2015.
Mahalanobis Distance

Squared Euclidean distance

\[ d(x_1, x_2) = \|x_1 - x_2\|_2^2 = (x_1 - x_2)^T(x_1 - x_2) \]

Let \( \Sigma = \sum_{i,j}(x_i - \mu)(x_j - \mu)^T \)

The Mahalanobis distance

\[ d_M(x_1, x_2) = (x_1 - x_2)^T\Sigma^{-1}(x_1 - x_2) \]
Metric Learning

Applying Mahalanobis distance to learn a positive semi-denite (PSD) matrix

\[ d_M(x_i, x_j) = \sqrt{(x_i - x_j)^T M (x_i - x_j)} \]

Relationship with subspace learning

\[
\begin{align*}
d_M(x_i, x_j) &= \sqrt{(x_i - x_j)^T M (x_i - x_j)} \\
&= \sqrt{(x_i - x_j)^T W^T W (x_i - x_j)} \\
&= \| W x_i - W x_j \|_2
\end{align*}
\]

where \( M = W^T W \).
Representative Metric Learning Algorithms

Large Margin Nearest Neighborhood (LMNN)

Minimize $\sum_{i,j} \eta_{ij} (\tilde{x}_i - \tilde{x}_j)^\top M(\tilde{x}_i - \tilde{x}_j) + c \sum_{i,j} \eta_{ij} (1 - y_{il}) \xi_{ijl}$ subject to:

1. $(\tilde{x}_i - \tilde{x}_l)^\top M(\tilde{x}_i - \tilde{x}_l) - (\tilde{x}_i - \tilde{x}_j)^\top M(\tilde{x}_i - \tilde{x}_j) \geq 1 - \xi_{ijl}$
2. $\xi_{ijl} \geq 0$
3. $M \succeq 0$.

[Weinberger et al, NIPS2005]
Representative Metric Learning Algorithms

Information-Theoretic Metric Learning (ITML)

\[
\min_A \quad \text{KL}(p(x; A_0) \| p(x; A))
\]

subject to \(d_A(x_i, x_j) \leq u \quad (i, j) \in S,\)
\(d_A(x_i, x_j) \geq \ell \quad (i, j) \in D.\)

where \(\text{KL}(p(x; A_0) \| p(x; A)) = \int p(x; A_0) \log \frac{p(x; A_0)}{p(x; A)} \, dx.\)

The optimization function can be re-formulated as

\[
\min_{A \geq 0} \quad D_{\ell_d}(A, A_0)
\]

s.t. \(\text{tr}(A(x_i - x_j)(x_i - x_j)^T) \leq u \quad (i, j) \in S,\)
\(\text{tr}(A(x_i - x_j)(x_i - x_j)^T) \geq \ell \quad (i, j) \in D,\)

[Davis et al, ICML2007]
Neighborhood Repulsed Metric Learning

Motivations

• For the verification task, the number of negative pairs is larger than the number of positive pairs if we know the exact label information of each sample.

• The importance of different negative pairs is different. Some negative pairs are very discriminative and some are not so discriminative.

• It is desirable to identity the most informative negative pairs and ignore the less informative negative pairs to learn a discriminative metric for verification.
Neighborhood Repulsed Metric Learning

\[
\max_A J(A) = J_1(A) + J_2(A) - J_3(A)
\]
\[
= \frac{1}{Nk} \sum_{i=1}^{N} \sum_{t_1=1}^{k} d^2(x_i, y_{it_1}) + \frac{1}{Nk} \sum_{i=1}^{N} \sum_{t_2=1}^{k} d^2(x_{it_2}, y_i)
\]
\[
- \frac{1}{N} \sum_{i=1}^{N} d^2(x_i, y_i)
\]
\[
= \frac{1}{Nk} \sum_{i=1}^{N} \sum_{t_1=1}^{k} (x_i - y_{it_1})^T A (x_i - y_{it_1})
\]
\[
+ \frac{1}{Nk} \sum_{i=1}^{N} \sum_{t_2=1}^{k} (x_{it_2} - y_i)^T A (x_{it_2} - y_i)
\]
\[
- \frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)^T A (x_i - y_i)
\]

Algorithm 1: NRML

**Input:** Training images: \( S = \{(x_i, y_i)|i = 1, 2, \cdots, N\} \),
Parameters: neighborhood size \( k \), iteration number \( T \), and convergence error \( \varepsilon \) (set as 0.0001).

**Output:** Distance metric \( W \).

**Step 1 (Initialization):**
Search the \( k \)-nearest neighbors for each \( x_i \) and \( y_i \) by using the conventional Euclidean metric.

**Step 2 (Local optimization):**
For \( r = 1, 2, \cdots, T \), repeat
2.1. Compute \( H_1 \), \( H_2 \) and \( H_3 \), respectively.
2.2. Solve the eigenvalue problem in Eq. (9).
2.3. Obtain \( W^r = [w_1, w_2, \cdots, w_l] \).
2.4. Update the \( k \)-nearest neighbors of \( x_i \) and \( y_i \) by \( W^r \).
2.5. If \( r > 2 \) and \( |W^r - W^{r-1}| < \varepsilon \), go to Step 3.

**Step 3 (Output distance metric):**
Output distance metric \( W = W^r \).

Multi-view Neighborhood Repulsed Metric Learning

\[ d_M(x_i, x_j) = \sum_{k=1}^{K} w_k^i (x_i^k - x_j^k)^T M_k (x_i^k - x_j^k) \]
Multi-view Neighborhood Repulsed Metric Learning

\[
\max_{W,\beta} \sum_{p=1}^{K} \beta_p \text{tr}[W^T (H_1^p + H_2^p - H_3^p)W]
\]

subject to \( W^T W = I, \sum_{p=1}^{K} \beta_p = 1, \beta_p \geq 0 \).

\[
\max_{W,\beta} \sum_{p=1}^{K} \beta_p^q \text{tr}[W^T (H_1^p + H_2^p - H_3^p)W]
\]

subject to \( W^T W = I, \sum_{p=1}^{K} \beta_p = 1, \beta_p \geq 0 \).

Algorithm 2: MNRML

**Input:** Training images: \( S^p = \{(x_i^p, y_i^p)\}_{i=1,2,\cdots,N} \) be the \( p \)th view set of \( N \) pairs of kinship images, Parameters: neighborhood size \( k \), iteration number \( T \), tuning parameter \( q \), and convergence error \( \varepsilon \) (set as 0.0001).

**Output:** Distance metric \( W \).

**Step 1 (Initialization):**

1.1. Set \( \beta = [1/K, 1/K, \cdots, 1/K] \);
1.2. Obtain \( W^0 \) by solving Eq. (18).

**Step 2 (Local optimization):**

For \( r = 1,2,\cdots,T \), repeat

2.1. Compute \( \beta \) by using Eq. (16).
2.2. Obtain \( W^r \) by solving Eq. (18).
2.3. If \( r > 2 \) and \(|W^r - W^{r-1}| < \varepsilon\), go to Step 3.

**Step 3 (Output distance metric):**

Output distance metric \( W = W^r \).
Experimental Results

• Comparisons with existing metric learning methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Feature</th>
<th>F-S</th>
<th>F-D</th>
<th>M-S</th>
<th>M-D</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSML</td>
<td>LBP</td>
<td>63.7</td>
<td>61.2</td>
<td>55.4</td>
<td>62.4</td>
<td>60.7</td>
</tr>
<tr>
<td></td>
<td>LE</td>
<td>61.1</td>
<td>58.1</td>
<td>60.9</td>
<td>70.0</td>
<td>62.5</td>
</tr>
<tr>
<td></td>
<td>SIFT</td>
<td>66.5</td>
<td>60.0</td>
<td>60.0</td>
<td>56.4</td>
<td>59.8</td>
</tr>
<tr>
<td></td>
<td>TPLBP</td>
<td>57.3</td>
<td>61.5</td>
<td>63.2</td>
<td>57.0</td>
<td>59.7</td>
</tr>
<tr>
<td>NCA</td>
<td>LBP</td>
<td>61.7</td>
<td>62.2</td>
<td>56.4</td>
<td>62.4</td>
<td>60.7</td>
</tr>
<tr>
<td></td>
<td>LE</td>
<td>62.1</td>
<td>57.1</td>
<td>61.9</td>
<td>69.0</td>
<td>62.3</td>
</tr>
<tr>
<td></td>
<td>SIFT</td>
<td>67.5</td>
<td>61.0</td>
<td>61.0</td>
<td>57.4</td>
<td>60.8</td>
</tr>
<tr>
<td></td>
<td>TPLBP</td>
<td>56.3</td>
<td>60.5</td>
<td>62.2</td>
<td>56.0</td>
<td>58.7</td>
</tr>
<tr>
<td>LMNN</td>
<td>LBP</td>
<td>62.7</td>
<td>63.2</td>
<td>57.4</td>
<td>63.4</td>
<td>61.7</td>
</tr>
<tr>
<td></td>
<td>LE</td>
<td>63.1</td>
<td>58.1</td>
<td>62.9</td>
<td>70.0</td>
<td>63.3</td>
</tr>
<tr>
<td></td>
<td>SIFT</td>
<td>69.5</td>
<td>63.0</td>
<td>63.0</td>
<td>59.4</td>
<td>62.8</td>
</tr>
<tr>
<td></td>
<td>TPLBP</td>
<td>57.3</td>
<td>61.5</td>
<td>63.2</td>
<td>57.0</td>
<td>59.7</td>
</tr>
<tr>
<td>NRML</td>
<td>LBP</td>
<td>64.7</td>
<td>65.2</td>
<td>59.4</td>
<td>65.4</td>
<td>63.7</td>
</tr>
<tr>
<td></td>
<td>LE</td>
<td>64.1</td>
<td>59.1</td>
<td>63.9</td>
<td>71.0</td>
<td>64.3</td>
</tr>
<tr>
<td></td>
<td>SIFT</td>
<td>70.5</td>
<td>64.0</td>
<td>64.0</td>
<td>60.4</td>
<td>63.8</td>
</tr>
<tr>
<td></td>
<td>TPLBP</td>
<td>59.3</td>
<td>63.5</td>
<td>65.2</td>
<td>60.0</td>
<td>62.9</td>
</tr>
<tr>
<td>MNRML</td>
<td>All</td>
<td>72.5</td>
<td>66.5</td>
<td>66.2</td>
<td>72.0</td>
<td>69.9</td>
</tr>
</tbody>
</table>

On KinFaceW-I dataset.

<table>
<thead>
<tr>
<th>Method</th>
<th>Feature</th>
<th>F-S</th>
<th>F-D</th>
<th>M-S</th>
<th>M-D</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSML</td>
<td>LBP</td>
<td>66.0</td>
<td>65.5</td>
<td>64.8</td>
<td>65.0</td>
<td>65.3</td>
</tr>
<tr>
<td></td>
<td>LE</td>
<td>71.8</td>
<td>68.1</td>
<td>73.8</td>
<td>74.0</td>
<td>71.9</td>
</tr>
<tr>
<td></td>
<td>SIFT</td>
<td>62.0</td>
<td>58.9</td>
<td>56.8</td>
<td>57.4</td>
<td>58.8</td>
</tr>
<tr>
<td></td>
<td>TPLBP</td>
<td>66.4</td>
<td>62.6</td>
<td>62.8</td>
<td>64.9</td>
<td>64.2</td>
</tr>
<tr>
<td>NCA</td>
<td>LBP</td>
<td>67.0</td>
<td>66.5</td>
<td>65.8</td>
<td>66.0</td>
<td>66.3</td>
</tr>
<tr>
<td></td>
<td>LE</td>
<td>73.8</td>
<td>70.1</td>
<td>74.8</td>
<td>75.0</td>
<td>73.5</td>
</tr>
<tr>
<td></td>
<td>SIFT</td>
<td>63.0</td>
<td>59.9</td>
<td>58.8</td>
<td>59.4</td>
<td>60.4</td>
</tr>
<tr>
<td></td>
<td>TPLBP</td>
<td>67.4</td>
<td>63.6</td>
<td>63.8</td>
<td>66.9</td>
<td>66.5</td>
</tr>
<tr>
<td>LMNN</td>
<td>LBP</td>
<td>68.0</td>
<td>68.5</td>
<td>68.8</td>
<td>67.0</td>
<td>68.2</td>
</tr>
<tr>
<td></td>
<td>LE</td>
<td>74.8</td>
<td>71.1</td>
<td>75.8</td>
<td>76.0</td>
<td>74.5</td>
</tr>
<tr>
<td></td>
<td>SIFT</td>
<td>65.0</td>
<td>57.9</td>
<td>58.8</td>
<td>59.4</td>
<td>60.4</td>
</tr>
<tr>
<td></td>
<td>TPLBP</td>
<td>68.4</td>
<td>65.6</td>
<td>65.8</td>
<td>67.9</td>
<td>68.1</td>
</tr>
<tr>
<td>NRML</td>
<td>LBP</td>
<td>69.0</td>
<td>69.5</td>
<td>69.8</td>
<td>69.0</td>
<td>69.5</td>
</tr>
<tr>
<td></td>
<td>LE</td>
<td>76.8</td>
<td>73.1</td>
<td>76.8</td>
<td>77.0</td>
<td>75.7</td>
</tr>
<tr>
<td></td>
<td>SIFT</td>
<td>68.0</td>
<td>60.9</td>
<td>60.8</td>
<td>61.4</td>
<td>62.8</td>
</tr>
<tr>
<td></td>
<td>TPLBP</td>
<td>70.4</td>
<td>67.6</td>
<td>67.8</td>
<td>69.9</td>
<td>70.1</td>
</tr>
<tr>
<td>MNRML</td>
<td>All</td>
<td>76.9</td>
<td>74.3</td>
<td>77.4</td>
<td>77.6</td>
<td>76.5</td>
</tr>
</tbody>
</table>

On KinFaceW-II dataset.
Experimental Results

• Comparisons with multi-view learning methods

<table>
<thead>
<tr>
<th>Method</th>
<th>KinFaceW-I</th>
<th>KinFaceW-II</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>68.4</td>
<td>74.3</td>
</tr>
<tr>
<td>MKL</td>
<td>67.5</td>
<td>73.4</td>
</tr>
<tr>
<td>MNRML</td>
<td>69.9</td>
<td>76.5</td>
</tr>
</tbody>
</table>

• Comparisons with human observers

Correct verification accuracy on the KinFaceW-I dataset.

<table>
<thead>
<tr>
<th>Method</th>
<th>F-S</th>
<th>F-D</th>
<th>M-S</th>
<th>M-D</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>HumanA</td>
<td>61.00</td>
<td>58.00</td>
<td>66.00</td>
<td>70.00</td>
<td>63.75</td>
</tr>
<tr>
<td>HumanB</td>
<td>67.00</td>
<td>65.00</td>
<td>75.00</td>
<td>77.00</td>
<td>71.00</td>
</tr>
</tbody>
</table>

Correct verification accuracy on the KinFaceW-II dataset.

<table>
<thead>
<tr>
<th>Method</th>
<th>F-S</th>
<th>F-D</th>
<th>M-S</th>
<th>M-D</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>HumanA</td>
<td>61.00</td>
<td>61.00</td>
<td>69.00</td>
<td>73.00</td>
<td>66.75</td>
</tr>
<tr>
<td>HumanB</td>
<td>70.00</td>
<td>68.00</td>
<td>78.00</td>
<td>80.00</td>
<td>74.00</td>
</tr>
</tbody>
</table>
Discriminative Multi-Metric Learning

$$\min_{M_1, \ldots, M_K, \alpha} J = \sum_{k=1}^{K} \alpha_k f_k(M_k) + \lambda g_k(W_1, \ldots, W_K)$$

subject to

$$\sum_{k=1}^{K} \alpha_k = 1, \alpha_k \geq 0.$$ where

$$f_k(M_k) = -\log(\prod_{o_1^k} P(g(x_i^k, y_j^k) < g(x_i^k, y_j^k)))$$

$$-\log(\prod_{o_2^k} P(g(x_i^k, y_j^k) < g(x_i^k, y_j^k)))$$

$$g_k(W_1, \ldots, W_K) = \sum_{k_1, k_2=1}^{K} \sum_{i=1}^{N} \|W_{k_1}^T x_i^{k_1} - W_{k_2}^T x_i^{k_2}\|_F^2$$

$$\min_{W_1, \ldots, W_K, \alpha} J = \sum_{k=1}^{K} \alpha_k f_k(W_k)$$

$$+ \lambda \sum_{k_1, k_2=1}^{K} \sum_{i=1}^{N} \|W_{k_1}^T x_i^{k_1} - W_{k_2}^T x_i^{k_2}\|_F^2$$

where

$$f_k(W_k) = \prod_{o_1^k} \log(1 + \exp(\|W_k^T x_i^{p} \|^2 - \|W_k^T x_i^{n} \|^2))$$

Discriminative Multi-Metric Learning

\[
\min_{W_k} J(W_k) = \alpha_k f_k(W_k) + \lambda \sum_{l=1, l \neq k}^K G(W_k)
\]

where

\[
G(W_k) = \sum_{i=1}^N \| W_k^T x_i^k - W_l^T x_i^l \|_2^2
\]

\[
\frac{\partial f_k(W_k)}{\partial W_k} = \prod_{O_k^l} \frac{2 + \exp(\| W_k^T x_{ik}^p \|_2^2 - \| W_k^T x_{ik}^m \|_2^2)}{1 + \exp(\| W_k^T x_{ik}^p \|_2^2 - \| W_k^T x_{ik}^m \|_2^2)} \times (x_{ik}^p x_{ik}^T - x_{ik}^n x_{ik}^n) W_k
\]

\[
\frac{\partial G(W_k)}{\partial W_k} = 2\lambda (K - 1) W_k \sum_{i=1}^N (x_i^k)^T x_i^k
\]

\[
-2\lambda W_k \sum_{l=1}^K \sum_{i=1}^N (x_i^l)^T x_i^l
\]

\[
W_k^{l+1} = W_k^l - \eta_k \frac{f_k(W_k)}{W_k} + \lambda \sum_{l=1, l \neq k}^K \frac{\partial G(W_k)}{\partial W_k}
\]

Iteration is terminated: \( J(W_k^l) - J(W_k^{l+1}) < \varepsilon \) or \( \| W_k^{l+1} - W_k^l \| < \varepsilon \)
Discriminative Multi-Metric Learning

\[
\min_{\alpha} \quad J(\alpha) = \sum_{k=1}^{K} \alpha_k^r f_k(W_k)
\]

subject to \[\sum_{k=1}^{K} \alpha_k = 1, \quad \alpha_k > 0.\]

The Lagrange function can be constructed as:

\[
L(\alpha, \zeta) = \sum_{k=1}^{K} \alpha_k^r f_k(W_k) - \zeta(\sum_{k=1}^{K} \alpha_k - 1)
\]

Let \(\frac{\partial L(\alpha, \zeta)}{\partial \alpha_k} = 0\) and \(\frac{\partial L(\alpha, \zeta)}{\partial \zeta} = 0\), we have

\[
ra_k^{r-1} f_k(W_k) - \zeta = 0
\]
\[
\sum_{k=1}^{K} \alpha_k - 1 = 0
\]

\[
\alpha_k = \frac{(\frac{1}{f_k(W_k)})^{1/(r-1)}}{\sum_{k=1}^{K} (\frac{1}{f_k(W_k)})^{1/(r-1)}}
\]
Prototype-Based Discriminative Feature Learning

\[
\text{max } H(B) = H_1(B) + H_2(B) - H_3(B)
\]
\[
= \frac{1}{Mk} \sum_{i=1}^{M} \sum_{t_1=1}^{k} \| f(x_i) - f(y_{i1}) \|_2^2
\]
\[
+ \frac{1}{Mk} \sum_{i=1}^{M} \sum_{t_2=1}^{k} \| f(x_{it_2}) - f(y_i) \|_2^2
\]
\[
- \frac{1}{M} \sum_{i=1}^{M} \| f(x_i) - f(y_i) \|_2^2
\]
subject to \( \| \beta_k \|_1 \leq \gamma, k = 1, 2, \cdots, K. \)

Summary and Future Work

- Metric learning is effective for kinship verification.
- Feature learning for kinship verification?
- Deep learning for kinship verification?