Soft-Margin Learning for Multiple Feature-Kernel Combinations with Domain Adaptation, for Recognition in Surveillance Face Dataset

Samik Banerjee* & Sukhendu Das#
VPLab, CS&E, IIT Madras, India
www.cse.iitm.ac.in/~vplab
*samik@cse.iitm.ac.in
#sdas@iitm.ac.in

CVPRW-Biometrics; June 2016
Overview

- Motivation and Problem definition
- Multiple-kernel learning
- Soft-Margin Learning for Multiple feature-kernel combination (SML-MFKC)
- Domain Adaptation (DA)
- Proposed Technology
- Results and discussion
- Conclusion & References
Motivation and Problem Statement

Problems:
• Low Resolution
• Blur
• Low contrast
• Aliasing effect
• Poor Illumination
• Different Camera Parameters

Aim:
Design of an effective Face Recognition algorithm, that exploits DA in RKHS to obtain a feature representation with similar distributions of gallery and probe, within an optimal feature-kernel combination by MKL.
Multiple Kernel Learning

• Given M kernel functions $K^1, ..., K^M$, find a positive combination of these kernels such that the resulting kernel $\mathcal{K}$ is « optimal » in some sense,

$$\mathcal{K}(x, x') = \sum_{m=1}^{M} w_m K^m(x, x'),$$

with $w \geq 0$, $\sum_{m} w_m = 1$

• Need to learn together the kernel coefficients $w_m$ and the SVM parameters.

• Extract features from all available sources

• Construct Kernel Matrices for
  ✓ Different Features
  ✓ Different Kernel Types
  ✓ Different Kernel Parameters

• How to find Optimal Kernel-Feature Combination and Kernel Classifier for each feature ?

Our aim is to transform the kernel-selection problem of MKL to a feature-selection* problem, to obtain an optimal feature-kernel combinations.

*(The method is modified based on the methodology proposed by Gehler et al., ICCV, 2009)*.
SML-MFKC

• Transformation of the kernel-selection problem of MKL to a feature-selection problem.

• Feature set: 
  \[ M = \{ \text{Eigenfaces, Fisherfaces, Weberfaces, Gaborfaces, BOW, FV-SIFT, VLAD-SIFT} \} \]

• Kernel Set: 
  \[ Q = \{ \text{Linear, Polynimial, Gaussian, RBF, Chi-square, RBF + Chi-square} \} \]

• Optimal feature-kernel combination \(< m^i, q^i >; m^i \in M; q^i \in Q. \)

• \( m^i \) is projected to RKHS using \( q^i \) to obtain a new feature representation.
SML-MFKC

• The cost function that gives the optimal combination of the kernel and the feature, \( q = 1, \ldots, P \):

\[
\sup \arg\min_{\beta^q \in \mathbb{R}^F} \frac{1}{2} \sum_{m=1}^F \beta_m^q \hat{\alpha}^T V_m^q \hat{\alpha} + C \sum_{i=1}^N L(y_i, b + \sum_{m=1}^F \beta_m^q \hat{\alpha}^T V_m^q \hat{\alpha}) \qquad \text{s.t. } \sum_{m=1}^F \beta_m^q = 1, \beta_m^q \geq 0, \forall m,
\]

where, \( L(y, t) = \max(0, 1 - yt) \),

\( P \) = No. of kernels, \( F \) = No. of features,

\( \beta_m^q \) = weight coefficient for the \( m \)-th feature \( (V_m^q) \) paired with the \( q \)-th kernel.

\( \hat{\alpha} \) and \( b \) are the parameters of SVM,

\( \beta^q \) is the local minima, \( C \) is a constant.

• Block-wise gradient-descent based approach is used for solving the minimization problem.

• The best selected feature, \( \tilde{V}_q \) is based on the supremum over the set, \( \beta^q \).
Domain Adaptation (DA)

Target Domain : Probe Samples
Source Domain : Gallery Samples

• Reasons for domain adaptation
  – Difference in resolution
  – Blur
  – Noise
  – Low-contrast
  – Different camera parameters

• Since the features in the RKHS will also have different distributions, we perform DA based on the technique proposed by Hoffman et al., ICLR, 2013.
Domain Adaptation (DA)

Proposed framework of DA:

- The normal to the affine hyperplane associated with the \( k \)-th kernel in a binary SVM is denoted as \( \theta_k; \ k = 1, \ldots, K \).
- The offset of that hyperplane from the origin is \( b_k \).
- \( W^T \) is the transformation of the source hyperplane parameters, \( \theta_k \).
- Training points & labels in the source domain:\{\( < x_1^S, y_1^S >, \ldots, < x_{n_S}^S, y_{n_S}^S > \)\}
- Target points & labels in the target domain:\{\( < x_1^T, y_1^T >, \ldots, < x_{n_T}^T, y_{n_T}^T > \)\}
- Hinge Loss: \( \mathcal{L}(y, x, \theta) = \max(0, 1 - \delta(y, k) \cdot x^T \theta) \)
- The cost function:

\[
J(W, \theta_k, b_k) = \frac{1}{2} \|W\|_F^2 + \sum_{k=1}^{K} \left[ \frac{1}{2} \|\theta_k\|_2^2 + C_S \sum_{i=1}^{n_S} \mathcal{L} \left(y_i^S, W \cdot \begin{bmatrix} x_i^S & [\theta_k] \end{bmatrix} \right) \right] + C_T \sum_{i=1}^{n_T} \mathcal{L} \left(y_i^T, \begin{bmatrix} x_i^T & [\theta_k] \end{bmatrix} \right)
\]

- \( C_S \) penalizes the source classification error.
- \( C_T \) penalizes the target adaptation error.
- The minimization of \( J \) using coordinate descent approach yields the transformation matrix \( W^T \).
Proposed Methodology

• Overall Framework

Datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>σ</th>
<th>γ</th>
</tr>
</thead>
<tbody>
<tr>
<td>FR_SURV</td>
<td>1.75</td>
<td>1.75</td>
</tr>
<tr>
<td>SCFace</td>
<td>1.70</td>
<td>1.70</td>
</tr>
<tr>
<td>ChokePoint</td>
<td>1.20</td>
<td>1.25</td>
</tr>
</tbody>
</table>
Training Phase

<table>
<thead>
<tr>
<th>Dataset</th>
<th>No. of Probe Samples for DA</th>
</tr>
</thead>
<tbody>
<tr>
<td>FR_SURV</td>
<td>5 per subject, from 20 random samples</td>
</tr>
<tr>
<td>SCFace</td>
<td>3 per subject, from 30 random samples</td>
</tr>
<tr>
<td>ChokePoint</td>
<td>6 per subject, from 5 males &amp; 2 females per profile</td>
</tr>
</tbody>
</table>
# Databases

<table>
<thead>
<tr>
<th></th>
<th>FR_SURV</th>
<th>SCFace</th>
<th>ChokePoint</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong># Subjects</strong></td>
<td>51</td>
<td>130</td>
<td>54</td>
</tr>
<tr>
<td><strong># Surveillance Cameras</strong></td>
<td>1</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Gallery Image size</td>
<td>150 x 150</td>
<td>250 x 250</td>
<td>80 x 80</td>
</tr>
<tr>
<td>Probe Image Size</td>
<td>33 x 33</td>
<td>15-45 x 15-45</td>
<td>80 x 80</td>
</tr>
<tr>
<td>Surveillance Scenario</td>
<td>Outdoor</td>
<td>Indoor</td>
<td>Indoor</td>
</tr>
</tbody>
</table>

**Gallery**

- D/B Name: FR_SURV
- Image: ![FR_SURV Gallery Image](image1)

**Probe**

- D/B Name: SCFace
- Image: ![SCFace Probe Images](image2)

- D/B Name: ChokePoint
- Image: ![ChokePoint Probe Images](image3)
## Experimental Results

### Rank-1 Recognition Rate (%)

<table>
<thead>
<tr>
<th>Sl.</th>
<th>Algorithm</th>
<th>SCFace</th>
<th>FR_SURV</th>
<th>ChokePoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EDA1 [14]</td>
<td>47.65</td>
<td>7.82</td>
<td>54.21</td>
</tr>
<tr>
<td>3</td>
<td>MDS [13]</td>
<td>42.26</td>
<td>12.06</td>
<td>52.13</td>
</tr>
<tr>
<td>4</td>
<td>KDA1 [14]</td>
<td>35.04</td>
<td>38.24</td>
<td>56.25</td>
</tr>
<tr>
<td>5</td>
<td>Gopalan [10]</td>
<td>2.06</td>
<td>2.06</td>
<td>58.62</td>
</tr>
<tr>
<td>6</td>
<td>Kliep [12]</td>
<td>37.51</td>
<td>28.79</td>
<td>63.28</td>
</tr>
<tr>
<td>7</td>
<td>Deep Face [15]*</td>
<td>41.25</td>
<td>29.35</td>
<td>62.15</td>
</tr>
<tr>
<td>8</td>
<td>Naïve</td>
<td>77.45</td>
<td>48.23</td>
<td>69.51</td>
</tr>
<tr>
<td>9</td>
<td>BaseMKL (only VLAD-SIFT)</td>
<td>53.36</td>
<td>36.54</td>
<td>66.12</td>
</tr>
<tr>
<td>10</td>
<td>Proposed</td>
<td><strong>79.86</strong></td>
<td><strong>56.44</strong></td>
<td><strong>85.59</strong></td>
</tr>
</tbody>
</table>

Comparison with several DL methods, on SURVEILLANCE Face Datasets

<table>
<thead>
<tr>
<th>No.</th>
<th>Method</th>
<th># CNN Layers</th>
<th>Rank-1 Recognition Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>FR_SURV</td>
</tr>
<tr>
<td>1</td>
<td>FV Faces + AlexNet</td>
<td>8</td>
<td>12.64</td>
</tr>
<tr>
<td>2</td>
<td>DeepFace [15]</td>
<td>19</td>
<td>29.35</td>
</tr>
<tr>
<td>4</td>
<td>DeepID-2,2+,3</td>
<td>60</td>
<td>34.94</td>
</tr>
<tr>
<td>5</td>
<td>FaceNet + Alignment [17]</td>
<td>22</td>
<td>36.53</td>
</tr>
<tr>
<td>6</td>
<td>VGG Face Descriptor + Deep Face [16]</td>
<td>16</td>
<td>32.57</td>
</tr>
<tr>
<td>7</td>
<td>Proposed Method</td>
<td>-</td>
<td>56.44</td>
</tr>
</tbody>
</table>
Experimental Results

• ROC Plots

• CMC Plots

FR_SURV   SCFace   ChokePoint
Conclusions

- An efficient method to tackle the problem of low-contrast and low-resolution proposed for near-frontal surveillance face images.
- Novel method using SML-MFKC to obtain an optimal pairing of feature and kernel, followed by DA in RKHS.
- Superiority of performance, shown using ROC, CMC and Rank-1 Recognition Rate than the other techniques, using three real-world surveillance face datasets.
- Deep learning fails to perform well for FR under surveillance scenarios.
- Larger set of subjects and performance with off-frontal faces may be studied, as future scope of work
References

References

THANK YOU!

sdas@iitm.ac.in

Queries?