Biometrics on Visual Preferences: a “Pump and Distill” Regression Approach

Cristina Segalin1, Alessandro Perina1, Marco Cristani2
1Computer Science Department, University of Verona, Verona, Italy
2Istituto Italiano di Tecnologia (IIT), Genova, Italy

Motivations and Goals

Behavioral biometric traits encode a characteristic linked to the behavior of a person [1]. The so-called HCI-based behavioral biometrics [5] are based on the idea that every person has a unique way to interact with a personal computer. We present a statistical behavioral biometric approach for recognizing people by their aesthetic preferences using color images, on a dataset of 40000 Flickr images.

The “Pump and Distill” Regression Approach

PHASE 0 – Features extraction:
Each image \( x \) is composed by the concatenation of features ranging from color statistics, image aesthetics cues, objects detections.

PHASE 1 – Pumping:
• Dataset divided in gallery and probe set (100 preferred images/user both)
• Cluster gallery set into \( K \) cluster \( \Rightarrow \) thematic exemplar \( \mu_k, k = 1, ..., K \)
• Augment the gallery set by bagging [3] \( \Rightarrow \) For each user \( u, v = 1, ..., V \), bagging generates \( G \) new training sets, the \( B_g^{(v)}, g = 1, ..., G \).

PHASE 2 – Distill:
Generate a set of surrogates \( \{ z \} \) for each bag \( B_g^{(v)} \), a user-customized representation of a cluster:
• Assign each image \( x \in B_g^{(v)} \) to the nearest thematic exemplar \( \mu_k \)
• Create a surrogate \( z \) by averaging all images associated to a given cluster

PHASE 3 – Regression:
Use surrogates of all the users to learn a per-user classifier performing LASSO [4]

Training:
• Assign to all the training surrogates \( \{ x_n \}, n = 1, ..., N^+ \) of a user the \( y_n = +1 \) label, and -1 to all the other \( N^- \) surrogates
• Regress a label assuming it as a linear combination of the image features: \( y_n = w^{(v^*)} z_n \) where \( w^{(v^*)} \) is the linear classifier for the user \( v \) obtained by minimizing the error function with the standard least squared estimate: \( E(w^{(v^*)}) = \sum_{n=1}^{N} (y_n - w^{(v^*)} z_n)^2 \)

Testing: Match a pool of probe images of user \( u \) with the gallery biometrical traits of user \( v \), represented by his positive surrogates

Results: Authentication

For every user \( u \), take the probe set of user \( u \) as client images, and all the other probe sets as impostor images.

Given an "authentication threshold", i.e. a value over which the subject is authenticated, sensitivity (TPR) and specificity (TNR) are computed using the averaged matching scores derived from the set of probe images. By varying this threshold, the ROC curve is obtained.

<table>
<thead>
<tr>
<th>( M = [v^*] N_u )</th>
<th>( \text{nAUC ROC} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M_{TPR=1} )</td>
<td>( 0.70 )</td>
</tr>
<tr>
<td>( M_{TPR=5} )</td>
<td>( 0.67 )</td>
</tr>
<tr>
<td>( M_{TPR=20} )</td>
<td>( 0.66 )</td>
</tr>
<tr>
<td>( M_{TPR=100} )</td>
<td>( 0.65 )</td>
</tr>
</tbody>
</table>

nAUC values of ROC curves varying num. of images per bag \( M \), and the \( M_{TPR} \).

Results: Recognition

Given the classifier template \( w^{(v)} \) of the user \( v \), the matching scores is aimed at measuring how likely the set \( \{ x^{(u)} \} \) of the user \( u \) contains images which are in accord with the surrogates \( \{ z^{(v)} \} \) by the user \( v \).

• Compute for every image \( x^{(u)} \) in the testing pool of user \( u \) the regression score \( \beta_m^{(u)} = w^{(v)} x^{(u)} \)

Final matching scores of the whole pool for each classifier \( w^{(v)} \) is determined as the averaged regression scores of the images belonging to it:

\[
\beta^{(u)}(v) = \frac{1}{M_{TPR}} \sum_{m=1}^{M_{TPR}} \beta_m^{(u)}
\]

• Build a CMC curve: given a probe set of images coming from a single user and the matching score, the curve tells the rate at which the correct user is found within the first \( r \) matches, with all possible \( r \) spanning on the x-axis (in our case, \( r = 1, ..., 200 \)).

CMC curve using \( M = 5 \) images per bag, and \( T_{E} = 20 \) test images, against the [2] approach. The highest differences in terms of recognition rate (the probability of having the correct match in the first \( r \) ranked positions) is localized in the first ranks, where is expected to find the correct match.

<table>
<thead>
<tr>
<th>( M = [v^*] N_u )</th>
<th>( \text{nAUC CMC} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M_{TPR=1} )</td>
<td>( 0.69 )</td>
</tr>
<tr>
<td>( 10 )</td>
<td>( 0.68 )</td>
</tr>
<tr>
<td>( 20 )</td>
<td>( 0.68 )</td>
</tr>
<tr>
<td>( 50 )</td>
<td>( 0.66 )</td>
</tr>
<tr>
<td>( 100 )</td>
<td>( 0.66 )</td>
</tr>
<tr>
<td>( [2] )</td>
<td>( 0.66 )</td>
</tr>
</tbody>
</table>

nAUC values of the CMC curves varying the num. of images per bag \( M \), and the num. of test images \( M_{TPR} \). We can think of [2] as having bags formed by one image. This suggests that pooling together images into surrogates produces more discriminative information.

References