Predicting trait impressions of faces using local face recognition techniques

Sheryl Brahnam a, Loris Nanni b, *  

a Computer Information Systems, Missouri State University, 901 S. National, Springfield, MO 65804, USA  
b DEIS, Università di Bologna, Viale Risorgimento 2, 40136 Bologna, Italy  

ARTICLE INFO  

Keywords:  
Trait impressions  
Face recognition  
Face classification  
Human–computer interaction  
Overgeneralization effects  
Classifier ensembles  

ABSTRACT  

The aim of this work is to propose a method for detecting the social meanings that people perceive in facial morphology using local face recognition techniques. Developing a reliable method to model people’s trait impressions of faces has theoretical value in psychology and human–computer interaction.  

The first step in creating our system was to develop a solid ground truth. For this purpose, we collected a set of faces that exhibit strong human consensus within the bipolar extremes of the following six trait categories: intelligence, maturity, warmth, sociality, dominance, and trustworthiness.  

In the studies reported in this paper, we compare the performance of global face recognition techniques with local methods applying different classification systems. We find that the best performance is obtained using local techniques, where support vector machines or Levenberg-Marquardt neural networks are used as stand-alone classifiers. System performance in each trait dimension is compared using the area under the ROC curve. Our results show that not only are our proposed learning methods capable of predicting the social impressions elicited by facial morphology but they are also in some cases able to outperform individual human performances.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction  

Over the last few decades, research in face recognition has expanded its focus to include some of the social aspects of faces, such as facial expression recognition. Emotional information is transitory in nature and is conveyed by deformations of facial features and by the dynamic movements involved in facial muscle contractions. Originally of theoretical interest, machine recognition of facial expressions is an established area of research, and we have witnessed rapid growth over the last decade in the application of this technology in such diverse areas as interface design, lip reading, synthetic face animation, human emotion analysis, face-image compression, security, and teleconferencing, to mention a few (Fasel & Luettin, 2003).

Another social aspect of faces that has sparked considerable interest in social psychology and other fields, but not so much in the area of machine classification, relates to the personality impressions that the static morphology of the face produces in the typical observer. Evidence indicates that people across cultures, age groups, genders, and social-economic status are consistent and similar to one another in their impressions of various facial forms (Zebrowitz, 1998). Several theoretical frameworks have been advanced that attempt to explain why certain facial characteristics consistently elicit specific impressions. One prominent approach is to examine facial appearances in terms of social affordances. In one variant of this theory, impression formation is based on a number of overgeneralization effects (McArthur & Baron, 1983). As outlined in Section 2, significant overgeneralization effects include facial attractiveness, maturity, gender, and emotion. According to this theory, specific facial characteristics advertize important information that lead people to behave in ways that increase their chances of survival. Recognizing a healthy potential mate, for example, increases the chances of producing robust offspring, and facial features indicative of health (rosy cheeks and unblemished skin textures) are generally sexually attractive. It is theorized that the positive effects of these facial characteristics are then overgeneralized so that attractive people, for instance, are assumed to possess a host of other desirable traits, such as intelligence, optimism, extraversion, and leadership and social skills.

In this paper, we attempt to classify faces according to the personality impressions they elicit. This task presents unique difficulties primarily because the ground truth is not based on ascertainable information about the subjects, such as identity, gender, and age. As noted in Sun, Sebe, and Lew (2004), developing a sound ground truth for detecting social aspects of faces is complicated even in the case of facial expression recognition, where research has identified a basic set of emotions that can be recognized using a well-established facial action coding system.
(Ekman & Friesen, 1978). Lacking a strong theoretical foundation for building a ground truth for the social impressions of facial morphology, our classes must be rooted in the impressions a set of faces make on the average observer. We are thus asking machines to map faces to a set of subjective social and psychological categories.

Few face classification experiments have focused on matching consensual subjective judgments. One early attempt is Brahnam (2002). In that study, principle component classifiers were trained to match the human classification of faces along the bipolar trait dimensions of adjustment, dominance, warmth, sociality, and trustworthiness. Results varied depending on the trait dimension; the classifiers matched the average observer in dominance (64%), adjustment (71%), sociality (70%), trustworthiness (81%) and warmth (89%). We believe, however, that these classification rates were inflated by the low number of faces (<15) in each trait class.

Related to our work are experiments designed to detect facial attractiveness. In Kanghae and Sornil (2007), for instance, two sets of photographs of 92 Caucasian females, approximately 18 years old, were rated by subjects using a 7-point scale. In their experiments, they classified faces into two classes, attractive (highest 25% rated images) and unattractive (lowest 25% rated images). Performance, based on percentage of correctly classified images, averaged 76.0% using K-nearest neighbor and 70.5% using linear support vector machines (SVMs). In this case, the authors believed that the low number of faces in the two classes (~24 images each) accounted for the poor performance of the classifiers.

As illustrated in Brahnam (2002) and Kanghae and Sornil (2007), it is important to develop a database that contains a sufficient number of representative faces within each trait class. In Section 3, we describe the databases first used in Brahnam and Nanni (2009) and in this study that attempt to circumvent some of the difficulties mentioned above in developing a solid ground truth for this problem domain. For the studies reported in Brahnam and Nanni (2009), six databases of faces were produced for the following six trait dimensions: intelligence, maturity, dominance, sociality, trustworthiness, and warmth. Each database was divided into two classes: high and low. Because artists were asked to design the faces, the number of faces verified by human subjects as belonging in each trait class averaged 111 faces, a much higher number of faces in each class than those used above. Single classifiers and ensembles were then trained to match the bipolar extremes of the faces in each of the six trait dimensions. With performance measured by the Area Under the ROC curve (AROC) and averaged across all six dimensions, results showed that single classifiers, especially linear SVM (0.74) and Levenberg-Marquardt neural networks (LMNNs) (0.73), performed as well as human raters (0.77). These single classifiers, however, performed poorly in the trait dimension of maturity. Ensembles of 100 LMNNs, constructed using Bagging (BA) (0.76), Random Subspace (SUB) (0.77), and Class Switching (CS) (0.75) compared equally well to rater performance but were better than the single classifiers at handling all six trait dimensions.

This study proposes using local methods for this classification task. It is well known in the biometric literature that methods based only on the global information of the images are not very effective with differences in facial expressions, illumination conditions, and poses. In the last few years, several face recognition methods based on local information have been proposed (Gottumukkal and Asari, 2004; Nanni and Maio, 2007; Shan et al., 2006; Tan and Chen, 2005). In Shan et al. (2006), the authors proposed using ensembles of classifiers, where the face image is first divided into M subimages (with all subimages having the same size without overlapping) according to their spatial locations. M feature vectors were then obtained by extracting the Local Binary Pattern (LBP) from the Gabor convolved subimage. A Fisher Discriminant Analysis (FDA) was then used to project the Gabor features onto a lower dimensional space. Finally, a nearest neighbor (NN) was performed in each space, and the results were combined.

LBP is a well known descriptor for textures and has been used for representing faces (Ojala, Pietikainen, & Maenpaa, 2002; Shan et al., 2006). In Shan et al. (2006) it is shown that if the LBP features are extracted from the Gabor convolved image, it is possible to obtain better performance with respect to that obtained by extracting the LBP features from the gray level image.

The remainder of this paper is organized as follows. In Section 2, we provide background information regarding this problem domain by summarizing the characteristics and traits associated with the overgeneralization effects of facial maturity, attractiveness, emotion, and gender. In Section 3, we present the study design and our method of generating the stimulus faces and of evaluating subject ratings. In Section 4, we describe our system architecture, and in Section 5 we compare the performance of simple classifiers and global classifier ensembles to the performance of different local methods. Our tests show that the local methods are more stable in performance across all six trait dimensions and closely approximate, if not exceed, the performance of individual human raters.

2. Social impressions of faces

In this section we provide some background material regarding the social impressions of facial morphology from the perspective of social psychology. As reviewed in Zebrowitz (2006), several theories have been advanced to explain why people form impressions of facial morphology that are fairly consistent. No one theory provides a complete understanding of this process. Several prominent social psychologists involved in facial impression research take an ecological approach to examining facial appearances in terms of their social affordances. As noted in the introduction, one important elaboration of the ecological approach is based on a number of overgeneralization hypothesis (McArthur & Baron, 1983). Research has shown that the structural characteristics of these effects are associated with specific sets of personality traits. As an introduction to this problem domain, we provide an overview in this section of the morphological characteristics and associated trait impressions of the following overgeneralization effects: facial attractiveness, maturity, gender, and emotion.

2.1. Overgeneralization effect of attractiveness

Attractive people are sexually desirable. Although debated, there is some evidence that attractiveness broadcasts actual indicators of physical, emotional, and intellectual fitness (Jones, 1995; Zebrowitz, Hall, & Murphy, 2002). The facial characteristics associated with attractiveness include proportion (Alley & Hildebrandt, 1988), symmetry (Grammer & Thornhill, 1994), straightness of profile (Carelo, Grososky, & Shaw, 1989), closeness to the population average (Langlois & Roggman, 1990), and sex specific features (Alley, 1992).

The overgeneralization effect of attractiveness hypothesizes that attractive people, being more sexually desirable, would be attributed more positive character traits. Much evidence supports what has come to be called the attractiveness halo effect. Attractive people are generally considered more socially adroit (Kuhlenschmidt & Croger, 1988), better leaders (Cherulnik, 1989), healthier, psychologically more adapted (Langlois, Kalakanis, & Rubenstein, 2000), intellectually more capable, and more moral than those less attractive. As would be expected, unattractive faces are associated with a host of negative traits, and they elicit more negative behaviors from people. Unattractive people are thought to be
untrustworthy (Bull & Rumsey, 1988), uncooperative (Mulford, Orbell, & Shatto, 1998), socially maladroit, unintelligent (Feingold, 1992), and psychologically unstable and antisocial (Langlois et al., 2000). People tend to gravitate towards the more attractive and ignore and avoid the unattractive (Bull & Rumsey, 1988). Attractive people are rewarded more often (Raza & Carpenter, 1987), whereas unattractive people suffer more inconsiderate behavior, aggression (Alcock, Solano, & Kayson, 1998) and abuse (Langlois et al., 2000).

2.2. Overgeneralization effect of babyfacedness

Infant faces trigger protective and nurturing behaviors in most mature animals and human beings (Lorenz, 1970–1971). The facial features that provoke these behaviors have been identified and include relatively large eyes, high eyebrows, red lips that are proportionally larger, and a small nose and chin. These facial features are also placed lower on the face (Berry & McArthur, 1985).

Given the strength of these nurturing-inducing facial stimuli, it is not surprising that adult faces that share some of the characteristics of baby faces also produce impressions similar to those associated with infants (Berry & McArthur, 1985). Babyfaced people are associated with common childlike characteristics as submissive-ness, naïveté, honesty, kind heartedness, weakness, and warmth (Berry & McArthur, 1985; McArthur & Apatow, 1984). Another strong correlation is lack of credibility and intellectual competence (Berry & McArthur, 1986). Babyfaced people are also thought to be more helpful, caring, and in need of protection (Berry & McArthur, 1986). In contrast, mature-faced individuals are more likely to command respect and to be perceived as experts (Zebrowitz, 1998). Facial maturity is characterized by a broader forehead, strong eyebrows, thin lips, lowered eyebrows, and narrow eyes (Keating, 1985).

2.3. Overgeneralization effect of gender

The trait impressions of babyfaced adults are similar to traits stereotypically attributed to females (Zebrowitz, 1998). This may be due to the fact that female faces tend to retain into adulthood more of the morphological characteristics of youth (Enlow & Hans, 1996). Female heads and facial features are smaller and the eyes tend to be more protrusive, making the eyes look larger. Female eyebrows are usually thinner and arched, resulting in eyebrows that look higher up on the face. In contrast, male faces tend have features associated with more mature faces: large nose, dominant jaw, angular cheekbones, deep-set eyes, and a more pronounced brow. The last two characteristics make the eyes look smaller (Enlow & Hans, 1996).

Females, like babyfaced individuals, are more likely to be ascribed characteristics associated with children. Women are stereotypically thought to be more submissive, caring, and in need of protection. Male faces, in contrast, are perceived as being more dominant, intelligent, and capable (Zebrowitz, 1998).

2.4. Overgeneralization effect of emotion

Research supports the idea that facial features that are associated with emotional states, such as smiling when happy, are regarded in ways appropriate to the emotional expression. People react positively to smiling faces and find them disarming and not very dominant (Keating, Mazur, & Segall, 1981). The overgeneralization effect of emotion predicts that a person whose lips resemble smiling would be perceived more positively. Research supports this hypothesis. People with upturned lips are considered friendly, kind, easygoing, and nonaggressive (Secord, Dukes, & Bevan, 1954). Similarly, facial morphology that resembles anger, i.e., faces that have low-lying eyebrows, thin lips, and withdrawn corners of the mouth, are perceived to be threatening, aggressive, and dominant (Keating et al., 1981).

3. Study design

In this section, we provide a brief description of the datasets used in our experiments. A complete description of the study design and datasets can be found in Brahm and Nanni (2009).

Our concern in developing the study design was to produce a set of faces that produced strong human consensus in the following trait categories: dominance, intelligence, maturity, sociality, trust-worthiness, and warmth. A slight modification of these categories was found to be most significant in the formation of person impressions according to a meta-analysis of the social psychology literature from 1967 to 1987 (Eagly, Ashmore, & Makhijan, 1991; Feingold, 1992; Rosenberg, 1977).

In Brahm and Nanni (2009), we justify our decision to explore machine classification of the social impressions faces make on the average human observer using artificially constructed faces, as opposed to photographs and 3-D scans. Each representation offers both benefits and drawbacks and can be compared with numerous social psychology studies. None of the above representations truly presents faces as they appear to us in our social environment (Brahm & Nanni, 2009). We chose to use facial composites because they have been widely studied in the person perception literature, and compared to 3-D scans are easier to collect.

In earlier studies, we randomized the construction of faces using the popular composite program FACES (Freierman, 2000). This produced a large number of faces that failed to produce significant impressions along the six trait dimensions. To remedy this problem, we had four artists design an equal number of male and female faces they thought fit each of the bipolar extremes of the trait dimensions. This resulted in the production of 480 stimulus faces that had a greater probability of eliciting strong trait impressions (see Fig. 1 for example faces). We then asked human subjects from a large Midwestern university to assess each face along each dimension using a 3-point scale. The midpoint was labeled neutral and the first and last points in the scale were randomly labeled, using polar opposite descriptors, e.g., intelligent/unintelligent and dominant/submissive. Subjects were provided with a definition of each trait descriptor.

A statistical analysis of the subject ratings produced the six datasets of faces listed in Table 1. Faces in each trait database were divided into two groups of low and high. For a face to be included in one of the six datasets, it needed to have a standard deviation less than 1.0, a mean rating <1.6 for low membership and >2.4 for high membership. Furthermore, the mode had to match the class (1 for low and 3 for high). Fig. 1 provides a sample of the faces that fell into the low (cold) and high (warm) classes for the trait dimension of warmth.

4. System architecture

Fig. 2 provides a basic schematic of the architecture of our proposed system. Starting from the idea that each subwindow (SW) provides unique information, we train different classifiers using the features extracted from a given SW. Moreover, we select only a subset of all the SWs by running Sequential Forward Floating Selection (SFFS) (Nanni & Lumini, 2007; Pudil, Novovicova, & Kittler, 1994).1 Hence, for each selected SW, a different classifier is trained, and then these classifiers are combined using the sum rule.

In the feature extraction step, the LBP features are extracted from the Gabor convolved image (Shan et al., 2006). The resultant Gabor feature set consists of the convolution results of an input SW image with 16 Gabor filters. Since we concatenate the invariant LBP (uniform patterns) histograms2 of 10 bins, 18 bins, and 26 bins, our feature vector is 864-dimensional \((864 = 16 \times (10 + 18 + 26))\) for each SW.

Gabor filters: The Gabor wavelets capture the properties of spatial localization, orientation selectivity, spatial frequency selectivity, and quadrature phase relationship (Shan et al., 2006). In the spatial domain, the 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave. The Gabor wavelet representation of an image is the convolution of the image with a family of Gabor kernels. Usually a number of Gabor filters of different scales and orientations are used. In this paper we apply Gabor wavelets at four different scales and four directions \((0^\circ, 45^\circ, 90^\circ, 180^\circ)\).

\(^2\) The Matlab implementation are available at the Outex site (http://www.ee.oulu.fi/mvg/page/lbp_matlab).

Table 1
Number of images in the two classes (high and low) of each trait dimension.

<table>
<thead>
<tr>
<th>Trait dimension</th>
<th>Attribution class</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intelligence</td>
<td>Low</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>140</td>
</tr>
<tr>
<td>Dominance</td>
<td>Low</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>102</td>
</tr>
<tr>
<td>Maturity</td>
<td>Low</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>126</td>
</tr>
<tr>
<td>Sociality</td>
<td>Low</td>
<td>126</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>117</td>
</tr>
<tr>
<td>Trustworthiness</td>
<td>Low</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>151</td>
</tr>
<tr>
<td>Warmth</td>
<td>Low</td>
<td>139</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>146</td>
</tr>
</tbody>
</table>

Note: Since a face can be rated as high or low in more than one trait dimension, the total number of faces in the six datasets is >480 (the total number of stimulus faces). Of course, within a given dimension, no face could be rated as both high and low.

In Ojala et al. (2002) the authors propose a theoretically and computationally simple approach, based on local binary patterns, which is robust in terms of grayscale variations and which is shown to discriminate a large range of rotated textures efficiently. Starting from the joint distribution of gray values of a circularly symmetric neighbor set of pixels in a local neighborhood, they derive an operator that is, by definition, invariant against any monotonic transformation of the grayscale. Rotation invariance is achieved by recognizing that this grayscale invariant operator incorporates a fixed set of rotation invariant patterns. A binary pattern \(LBP_{\text{GR}}\) is calculated by considering the difference between the gray value of the pixel \(x\) from the gray values of the circularly symmetric neighborhood set of \(P\) pixels placed on a circle radius of \(R\); when a neighbor does not fall exactly in the center of a pixel, its value is obtained by interpolation. The LBP histogram of dimension \(n\) (usually \(n = P + 2\)) is obtained by considering all the binary patterns of a given image. In this work we have tested the three configurations proposed in Ojala et al. (2002), namely, \((P = 8; R = 1; n = 10), (P = 16; R = 2; n = 18), (P = 24; R = 3; n = 26)\).

Sequential Forward Floating Selection\(^1\) (SFFS) is a top down search that successively deletes features from a set of original candidate features in order to find a smaller optimal set of features. As the number of subsets to be considered grows exponentially with the number of original features, these algorithms provide a heuristic for determining the best order to transverse the feature subset space. The best feature subset, \(S_k\) of size \(k\), is constructed by adding a single feature to the subset, \(S_{k-1}\) (with \(k = 1\) initially equal to 0), that gives the best performance for the new subset. At this point, each feature in \(S_k\) is excluded, and the new set \(S_{k-1}\) is compared with \(S_{k-1}\). If \(S_{k-1}\) outperforms \(S_{k-1}\), then it replaces \(S_{k-1}\).

We use SFFS as a feature selection method, where each feature, as noted above, is a given SW (i.e., a selection of the face). We do this to find the most useful SWs for classification in the whole face. Therefore, we select (one for all) the classifiers to be combined by SFFS, using the minimization of the Area Under the ROC curve (AROC) as the objective function (see Section 4).
Given our proposed architecture, several methods can be built using different classifiers and methods for SW positioning. In this paper, we propose three different systems:

1. SW30(k), where each image is divided into 28 SWs (size $30 \times 30$ pixels) and where the LBP features are extracted and convolved with Gabor Filters from the SWs (Shan et al., 2006). Linear SVMs are used to classify the data. We select (one for all) the best $k$ classifiers to be combined by running SFFS using AROC as the objective function on each of the six datasets. For each dataset we randomly resample the learning and the testing sets (containing respectively half of the patterns) twenty times.

2. SW20(k), as in 1 above, except in this method each image is divided into 60 SWs of dimension $20 \times 20$ pixels.

3. SW20RS(k), as in 2 above, except in this method the classifier we use is a random subspace of an ensemble of 100 LMNNs with 5 hidden nodes. To reduce dimensionality, the Gabor + LBP features are projected onto a lower PCA subspace (where the preserved variance is 0.98) before training the ensemble.

5. Classification experiments

We compare the performance gained by our local methods with the following four different global methods (as proposed in Brahnam & Nanni (in press)):

1. PCA + NN, classifier based on the PCA coefficients extracted from the gray values (where the preserved variance is 0.98) using the Euclidean distance;

2. PCA + SUB, classifier based on the PCA coefficients and Oja’s subspace maps, where subspaces are computed for each class of the dataset and the distance of the test patterns to these subspaces are used to classify the data;
3. PCA + SVM, classifier based on the PCA coefficients and a Linear SVM classifier (Vapnik, 1995);
4. LEM + SVM, classifier based on the Laplacian eigenface (50 dimensions are retained and preprocessed using PCA with preserved variance of 0.98 (He, Yan, & Hu, 2005)) and a linear SVM.

In addition, we compare these systems with standard methods for building ensembles of classifiers, as reported in Brahnam and Nanni (2009). We selected the LMNN with 5 hidden nodes as the classifier for building the ensembles since it is well known that ensembles work better if unstable classifiers are used (Kuncheva & Whitaker, 2003). The ensembles were constructed using the following three methods (see Brahnam & Nanni (2009) for a more detailed description of each of the following methods):

1. PCA + RS, classifier based on the PCA coefficients and a random subspace (RS) ensemble of 100 LMNNs. In RS the individual classifiers use only a subset of all features for training and testing (Ho, 1998). The percentage of the features retained in each training set was set to 50%.
2. PCA + BA, classifier based on the PCA coefficients and a bagging (BA) ensemble of 100 LMNN. Given a training set \( S \), BA generates \( K \) new training sets \( S_1, \ldots, S_K \); each new set \( S_i \) is used to train exactly one classifier (Breiman, 1996). Hence an ensemble of individual classifiers is obtained from \( K \) new training sets.
3. PCA + CS, classifier based on the PCA coefficients and a class switching (CS) ensemble of 100 LMNN. CS creates an ensemble that combines the decisions of classifiers generated by using perturbed versions of the training set where the classes of the training examples are randomly switched (Martínez-Muñoz & Suárez, 2005).

6. Results

As previously noted, the performance indicator adopted in this work is the AROC (Ling, Huang, & Zhang, 2003). AROC is a scalar measure to evaluate performance which can be interpreted as the probability that the classifier will assign a higher score to a randomly picked true positive sample than to a randomly picked true negative sample.

In this section, we report the average AROC obtained randomly by resampling 20 times the learning and the testing sets (each set containing half of the patterns). In Fig. 3, we compare the performance gained using the different global methods with our local methods (where a stand-alone classifier is trained for each SW). In Fig. 4, we report the average AROC obtained in the six trait problems studied in this work.

The tests reported in Fig. 5 are aimed at comparing the verification performance gained by our system as a function of the number, k, SWs selected by SFFS. We plot the AROC obtained combining the k best matchers selected by SFFS. In Fig. 5a, we report the performance of the SW30 method and compare that in Fig. 5b to the performance of the SW20 method.

The conclusions that may be extracted from the results reported in Figs. 3–5 are the following. First, we see that local approaches with SW selection dramatically outperform global approaches. As can be seen, without feature selection, the performance of the local approaches is similar to the best global methods. This indicates to us that some SWs are very useful in this classification problem, whereas others are not. Second, we note that both the SW20 and the SW30 methods, with \( k = 3 \), obtains a performance similar to the best global approach. The best SWs selected in the SW20 and SW30 methods are shown in Fig. 6, and the worst SWs are shown in Fig. 7. Classifying the worst SWs using SW30, for example, obtained an average AROC of only 0.55.

In Fig. 8, we compare the performance gained using the standard ensemble of classifier methods, described in this section above, and our local methods, described in Section 4, where an ensemble of classifiers is trained for each SW.

From this test it is clear that a global method based on an ensemble of LMNNs outperforms a global method based on SVM. Among the methods for building an ensemble of classifiers, the RS is the best method. The local method SW20RS and the local method SW20 obtains a very similar result, with both obtaining an AROC curve ~0.82.

In addition to the experiments reported above, we ran SW20 (the system with the best trade-off between performance and computation time), using a different protocol to select the best SWs. For each of the six trait datasets, the used SWs are selected considering the other five datasets. For instance, the SWs used to classify the trait category of intelligence are selected considering the trait categories of maturity, warmth, sociality, dominance, and trustworthiness. We find that using this blind testing protocol the AROC curve is very high (~0.80).

It is important to stress that trait impressions are subjective and that two different people might assign different trait scores to the same face. We found that the AROC of human raters to be 0.77. To calculate this value, we use the score assigned to each image from each user to calculate his individual performance. These individual performances were then averaged. It is interesting to note that the performance obtained by a local ensemble is higher than the average performance of the individual raters. This result leads us to believe that machines are as capable of classifying faces according to the impressions they make on the general observer as are most human beings.

7. Conclusion

This study indicates that machines are capable of perceiving faces in terms of the social impressions they make on people. Using a dataset we specifically developed for this task, we trained single classifiers and ensembles to match human impressions of faces in the bipolar extremes of the following trait dimensions: intelligence, maturity, warmth, sociality, dominance, and trustworthiness.

Two single classifiers, SVMs and LMNN, were able to closely match the performance of human raters at the task. However, local face recognition techniques proved superior to the single classifiers.
in their ability to handle all six dimensions. Our local methods even outperformed human raters in the averaged AROC of the six trait dimensions.

In Section 3 we noted a number of limitations using various methods of representing faces (e.g., photographs, facial constructions, and 3-D scans). The faces in this study were constructed from the database of facial features in the popular composite program FACES. Four artists were asked to produce 480 stimulus faces with an eye towards making faces they thought were clearly intelligent, unintelligent, mature, immature, warm, cold, social, unsocial, dominant, submissive, trustworthy, and untrustworthy. This method of generating faces succeeded in producing an average of 111 faces that were verified to elicit the impressions associated with the above classes. For future work, we are developing datasets of photographs of natural faces that have equally large numbers of faces in each of the twelve trait classes.

At present, our interest in building systems that match human classification of faces in terms of the personality impressions they make is purely theoretical. However, as with emotion recognition, we anticipate the development of application areas. We believe our local results and the SWs we have isolated could help social psychologist investigate further the specific facial characteristics that lead people to make various trait attributions. Aside from developing systems and models that may prove valuable in social psychology, research in this area could also prove useful in human–computer interaction (HCI). Many HCI researchers anticipate that human-like interfaces and robots of the future will need to possess an understanding of the social value of objects, including faces, if they are to engage with us in socially meaningful and gratifying ways. It is not enough in human society simply to recognize what an object is; one is expected to be able to manipulate objects in terms of the cultural layers of meanings that envelop them. Our preliminary research in this area indicates that it is possible for machines to detect some of the social significance embedded in images.

References