AUTOMATIC ANALYSIS OF FACIAL ATTRACTIVENESS FROM VIDEO

Sacide Kalaycı, Hazım Kemal Ekenel

Istanbul Technical University
Faculty of Computer and Informatics
{kalaycisa,ekenel}@itu.edu.tr

Hatice Gunes

Queen Mary University of London
School of EECS
h.gunes@qmul.ac.uk

ABSTRACT

There has been a growing interest in the computer science field for automatic analysis and recognition of facial beauty and attractiveness. Most of the proposed studies attempt to model and predict facial attractiveness using a single static facial image. While a static image provides limited information about facial attractiveness, using a video clip that contains information about the motion and the dynamic behaviour of the face provides a richer understanding and valuable insights into analysing facial attractiveness. With this motivation, we propose to use dynamic features obtained from video clips along with static features obtained from static frames for automatic analysis of facial attractiveness. Support Vector Machine (SVM) and Random Forest (RF) are utilised to create and train models of attractiveness using the features extracted. Experimental results show that combining static and dynamic features improve performance over using either of these feature sets alone, and SVM provides the best prediction performance.

Index Terms— Facial attractiveness, automatic analysis, static and dynamic features.

1. INTRODUCTION

Human beauty or attractiveness is a dominant aspect (trait) of the physical personality [1]. When we meet others, we attempt to incorporate information about them into our framework of intentions, motives, and causal relations, all of which are used to compose a model of the other and can be used to predict how the other may act. Attractiveness therefore plays an important role in meeting people, and in determining which social relationships will be pursued [2]. Contemporary research findings on attitudes and personality revealed that people respond positively to attractiveness and associate it with positive character traits (e.g., socially competent, potent, intellectually capable and psychologically more adapted) [1]. These in turn bring to mind the question of what (morphological) characteristics make a human (face / body / voice) attractive? Are there any universally accepted and recognized aesthetic canons? Ancient philosophers believed that there is a timeless, aesthetic ideal concept of beauty based on proportions, symmetry, harmony, and geometry, independent of the observers themselves. The Renaissance artists popularized this view by formulating the ideal proportions of the human form (e.g., ideal facial proportions were defined with the theory of equal thirds). Until the nineteenth century these views had a major influence on the Western perception of beauty and use of certain artistic canons. In the twentieth century anthropometrists started challenging these claims by conducting a number of experiments (e.g., [3],[4]). Modern view of beauty thus has been based on the idea that it is variable and subjective to race, culture or era. Faces that are more familiar are considered more attractive [5], and there is evidence suggesting that moral judgments influence opinions of beauty.

Research on quantifying and computing beauty and attractiveness has predominantly focused on analyzing the face. Aarabi et al. [6] introduced an automatic beauty analyzer that extracts geometric ratios of distances between a number of facial feature points and uses k-nearest neighbors (k-NN) to classify facial images into one of the four beauty categories (91% correct classification on a validation set of 40 images). The beauty predictor of White et al. [7] uses textural features to predict the mean attractiveness scores assigned to 4000 face images (downloaded from `www.hotornot.com`) using ridge regression (with a Gaussian RBF kernel). The best prediction results (a correlation coefficient of 0.37) were obtained using kernel Principal Component Analysis (PCA) on the face pixels. The system of [8] analyzes frontal facial images in terms of golden proportions and facial thirds by representing each face in terms of distances between facial features. The standardized classifier error (by using variance in human ratings) was found to be on average less than the standard deviation within the class. Eisenthal et al. [9] focused on classifying face images as either attractive or unattractive using SVMs, k-NN, and standard linear regression. When tested on two databases each containing 92 images of young women with neutral facial expressions, best results were obtained using geometric features based on pairwise distances between fiducial points (a correlation coefficient of 0.6) using linear regression and SVMs (eigenface projections provided a correlation coefficient of 0.45).

The attractiveness predictor of Kagian et al. [10] uses 90 principal components of 6972 distance vectors (between 84

978-1-4799-5751-4/14/$31.00 ©2014 IEEE

fiducial point locations) and standard linear regression to predict mean attractiveness scores of female facial images. When tested on the female Israeli database of [9], they achieved a correlation of 0.82 with mean attractiveness scores provided by human raters (along a range 1–7). Whitehill and Movellan [11] presented an automatic approach to learning the personal facial attractiveness preferences using a variety of low level representations such as PCA, Gabor filter banks, and Gaussian RBFs as well as image representations based on higher-level features (i.e., automated analysis of facial expressions, and SVMs for regression. When evaluated on a dataset of images collected from an online dating site, the system achieves correlations of up to 0.45 on the attractiveness predictions for individual users. Chen and Zang introduced a benchmark database for (female and male) facial beauty analysis in [12], and extracted geometric features. Their results showed that averageness hypothesis and symmetry hypothesis reveal much less beauty related information than multivariate Gaussian model. Sutic et al. [13] chose to combine eigenface and ratio-based feature representation and compared k-NN, neural network and AdaBoost algorithms for a two-class (more vs. less attractive) and a four-class (with quartile class boundaries: 3.0, 7.9, and 9.0 of maximum 10) attractiveness classification problem on a dataset of 2250 female images (extracted from the website `www.hotornot.com`). For the two-class problem, 61% classification accuracy was obtained using k-NN and geometric features, and 67% classification accuracy was obtained using k-NN and the distances in the eigenface space.

As indicated by the reviewed works, most of the proposed studies attempt to analyse facial attractiveness using a static facial image. While a static image provides limited information about facial attractiveness, using a video clip that contains information about the motion and the dynamic behaviour of the face provides a richer understanding and a fuller insight into modelling and analysing facial beauty. With this motivation, we propose to use dynamic features obtained from video clips along with static features obtained from static frames for automatic analysis of facial attractiveness. Subsequently, two different attractiveness prediction models are trained using Random Forest [14] and Support Vector Machines [15]. Comparative experiments show that combining static and dynamic features improve performance over using either of these feature sets alone, and SVM provides the best prediction performance.

The rest of the paper is organized as follows. Section 2 discusses extraction of dynamic and static features. Section 3 gives an overview of the dataset. Section 4 describes the training and the testing process. Section 5 presents experimental results. Finally section 6 concludes the paper.

2. FEATURE EXTRACTION

Studies in the literature suggest that the best recognition results for analysing facial attractiveness have been obtained by using features based on computing facial proportions, and distances and angles between facial features [9, 10]. Therefore, we utilise the Supervised Descent Method (SDM) [16] to locate and track 49 facial landmarks for each frame in a video. The coordinates of these 49 points are used to calculate geometric features as shown in Figure 1. Using these landmarks, we calculate static and dynamic feature characteristics. Deriving static features enable us to analyse geometric properties that are informative of facial beauty, namely symmetry, big eyes, thick lips etc. Dynamic features instead provide us information about the facial behaviour of the person, e.g., how expressive their face is, how they move their facial features etc.

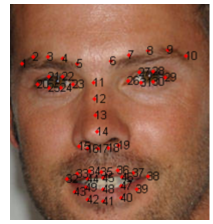

Fig. 1: 49 features localized using the SDM method.

The distances measured between facial points may vary depending on the distance between the face and the camera. To mitigate for this issue, the computed distances between feature points are normalized. To normalize for the facial height, the distance between point 5,6 (p5,p6) and the middle point between points 34,36 (p34,p36) and points 40,42 (p40,p42) are used. To normalize for the facial width, the mean of the distance between point 20 (p20) and point 29 (p29), and the distance between point 1 (p1) and point 10 (p10) are used.

Static Features. Previous works demonstrated that the most important features that are informative for automatic prediction of attractiveness are symmetry and the similarity between golden ratios, and the facial third or facial fifth ratios [17]. Based on these findings, we obtain the static features using the first frame of a video clip, and estimate ratios that are informative of facial symmetry and facial golden ratios. We extracted 15 static features in total which are listed in Table 1.

Mouth width to interocular distance	(d(p(32),p(38))/d(p(23),p(26)))
Eye fissure width to eye height	d(p(29),p(26))/(d(p(27),p(31))+d(p(28),p(30))/2)
Interocular distance to eye fissure width	d(p(23),p(26))/d(p(29),p(26)))
Lip height to lip width	d(p(35),p(41))/d(p(32),p(38))
Interocular distance to lip height	d(p(23),p(26))/d(p(35),p(41)
Lip height to nosemouth distance	d(p(35),p(41))/d(p(17),p(35))
Eye fissure width to nosemouth distance	d(p(29),p(26))/d(p(17),p(35))
Facial symmetry (8 features)	(5,6),(1,10),(2,9),(15,19),(20,29),(23,26) (32,38),(34,36) symmetry ratios of points

Table 1: List of the static features extracted.

Dynamic Features. Although studies on facial attractiveness to date give us an opinion on what kind of static features contribute to the perception of facial attractiveness, there is no study on the effects of dynamic features for the perception and judgement of facial beauty. As detailed information about what kind of dynamic features could have effect on facial attractiveness perception is lacking, we gathered knowledge for determining the dynamic features to be extracted from other domains, such as the field of automatic facial expression recognition [18]. In total 21 dynamic features are extracted by calculating features such as movement of the eye-brow and movement of the lips. The full set of dynamic features extracted are shown in Table 2.

Head movements (3 features)	Mean position of right-brow
Mean position of the head (3 features)	Mean position of left-brow
Position of lip corners	Movements of right-brow
Movements of lips (4 features)	Movements of left-brow
Number of apices	Eye aperture (2 features)
Movements of eyelid (2 features)	Total length of apices

Table 2: List of the dynamic features extracted.

3. DATASET

Although there are a number of datasets for analyzing facial attractiveness [6, 10, 15], none of them is suitable for the purpose of our study due to the fact that virtually all of them contain static facial images only. Therefore we used a subset from the SEMAINE dataset [19] which has been recorded to study the behavioral changes and different affect manifestations by a user interacting with four *virtual characters*, each with a distinct emotional style, and a conversational goal of shifting the user towards that state. These four characters are Prudence, even-tempered and sensible; Poppy, happy and outgoing; Spike, angry and confrontational; and Obadiah, sad and depressive. The goal is to elicit different types of user emotional and social behavior while user interacts with these virtual characters. Our subset contains 45 clips in total extracted from the SEMAINE database. These clips consist of visual recordings of 11 different participants interacting with the four aforementioned characters. Each clip was assessed for facial attractiveness by 6 raters and scored on a Likert scale with ten possible values, from *strongly disagree* to *strongly agree*, mapped into the range from [1,10]. The extracted clips from the SEMAINE database are curtailed on average to 14.09s with recordings containing 45.5% male and 54.5% female users. Since one subject's face appears to be largely occluded, the 4 video clips have been removed, leaving us 41 video clips in total. Sample frames from these video clips are shown in Figure 2.

For each video clip, we calculated the mean attractiveness scores across all 6 raters as ground truth attractiveness scores. Using only the mean score as ground truth may be misleading as there are cases where although attractiveness scores for video clips are different, the calculated mean score is very similar [20]. In order to mitigate for this issue, standard deviation is also calculated for each video clip along with the mean score.

Fig. 2: Sample frames from the data set used for the study.

4. ANALYSIS OF FACIAL ATTRACTIVENESS

After 36 features are extracted for each video,a feature vector with 41*36 dimension is created and scaled to [-1, +1]. A feature selection strategy is then applied to obtain a second feature set with the selected features only.

Instead of classifying attractiveness into fixed number of categories, we use regression methods for predicting the levels of facial attractiveness. We suggest that the attractiveness dimension should be considered as a continuum, and providing continuous predictions along this dimension will be more descriptive for evaluating attractiveness. To this aim, we utilise Support Vector Regression (SVR) and Random Forest Regression (RFR). Best performance was obtained with $g = 2$ and $s = 8$ for SVR, and $numberoftrees = 500$ for RFR. To benefit from as many training samples as possible during training and evaluation, we chose the leave-one-video-out cross-validation method. As we are aiming to produce continuous predictions along [1,10], we adopt the cross-correlation and root mean square error metrics for evaluating the performance of the system. Correlation (COR) is calculated using the equation (1) where x_i and y_i correspond to predicted scores of i^{th} the video and the ground truth score, respectively, and $\overline{x}$ and $\overline{y}$ are identified as mean value of these scores.

$$COR(X,Y) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{(\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2)^{1/2}} \quad (1)$$

To calculate the root mean square error (RMSE) equation (2) is used, where y_t and $\overline{y_t}$ correspond to predicted scores of t^{th} and ground truth score, respectively.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \overline{y_t})^2}{n}} \quad (2)$$

5. EXPERIMENTAL RESULTS

Feature Selection. As has been shown by existing literature, not all extracted features are relevant for automatic analysis of facial attractiveness. Irrelevant feature elimination reduces the computational complexity of the learning algorithms, by reducing the level of noise and provides better prediction performance. Along with these advantages, it also enables

978-1-4799-5751-4/14/$31.00 ©2014 IEEE

us to be aware of which features play a major role for automatic prediction of facial attractiveness. The most common approach to identify relevant and informative features is to assign scores to subsets of features based on proven performance, and to use searching algorithms to find the subset which provides the highest score. However, searching the whole set is virtually impossible as the large number of features generates a vast number of subsets. Evaluating and scoring relevant features separately is another option. However, evaluating individual features separately eliminates the relationships that might exist between groups of features, which when used together may provide better results. To avoid this undesirable situation, Navot & Shpigelman presented a method entitled RGS (Regression,gradient guided,feature selection) [21]. In this method, mean square error is calculated and features are scored using k-nearest neighbor algorithm and LOO(Leave-one-out) method. Moreover, RGS considers the weight of all features simultaneously, and therefore it can handle dependency on the groups of features. In our work, 36 features are ranked using the RGS method, and 10 features with the highest score are determined. The selected features are shown at Table 3 (s= static feature; d= dynamic feature). On closer inspection, in the list of best 10 features selected, although the most important features appear to be the static ones, there are 3 static and 7 dynamic features in total. This indicates that dynamic facial features play a significant role in predicting facial attractiveness.

1(s)	Interocular distance to eye fissure width
2(d)	Mean position of left-brow
3(d)	Number of apex
4(d)	Head movement
5(d)	Mean position of right-brow
6(d)	Eye aperture
7(d)	Movements of lips
8(d)	Movements of eyelid
9(s)	Interocular distance to lip height
10(s)	Lip height to nosemouth distance

Table 3: The best 10 features selected by RGS.

All Features vs. Best-10 Features. We wanted to evaluate how prediction performance depends on selecting the best 10 features versus using the full feature set. To this aim, we utilise again SVR and RFR methods, using the two feature vectors with $41 * 36$ features and with $41 * 10$ features, respectively. The results are shown in Table 4. The results clearly indicate that SVR method that uses the top 10 features provides the best prediction performance. In addition, eliminating irrelevant features contribute to improving the performance of both machine learning techniques. Since SVR method provided the best prediction performance, we further compare the prediction errors of SVR with the standard deviation of the observer scores. We conclude that SVR makes high prediction errors on video clips which have high standard deviations among the observer scores. This is in line with our earlier statement that mean value of the observer scores may not be sufficient as the ground truth value to describe the level of facial attractiveness.

		SVR	RFR
36 Features	COR	0,76	0,72
	RMSE	0,73	0,82
10 Features	COR	0,79	0,73
	RMSE	0,70	0,81

Table 4: Experimental results using SVR and RFR.

		Static	Dynamic	*Static + Dynamic*
SVR	COR	0,61	0,39	0,76
	RMSE	0,89	1,17	0,73

Table 5: Experimental results using static and dynamic features individually or together.

Static Features vs. Dynamic Features. In order to further understand whether dynamic features or static features play a more important role, we divided 36 features into two groups composed of 15 static and 21 dynamic features. Results are obtained by feeding these two feature vectors into SVR, and these are shown in Table 5. The table shows that facial attractiveness prediction is improved when static and dynamic features are used together.

6. CONCLUSION

We introduced a novel approach to automatic attractiveness prediction by analysing dynamic facial features along with the static ones. To the best of our knowledge, dynamic facial features have not been explored for beauty analysis and prediction to date. We obtain static and dynamic sets of features using the SDM method that detects and tracks facial feature points. These feature sets are then fed into SVR and RFR for automatic prediction of facial attractiveness along a continuum of 1 to 10. Experimental results show that using the top 10 static+dynamic features with SVR provides the best prediction performance. When we use the static and dynamic features separately for prediction, results show that combining the static and dynamic features improves performance over using either of these feature sets alone. Our results show that in order to fully understand the perception of facial attractiveness, the dynamics of facial behaviour need to be investigated further along with appearance features such as skin texture and eye/lip color.

7. ACKNOWLEDGMENTS

This work is partially supported by the British Council under the UK-Turkey Higher Education Partnership Programme (Grant Ref: TR/012012/KP40). Hatice Gunes work is supported by the EPSRC MAPTRAITS Project (Grant Ref: EP/K017500/1). The work is also supported by TUBITAK within the CHIST-ERA project titled CAMOMILE, project no. 112E176; and a Marie Curie FP7 Integration Grant within the 7th EU Framework Programme.

8. REFERENCES

[1] S. Branham, "Creating physical personalities for agents with faces: Modeling trait impressions of the face," in *Proc. of Workshop on Attitudes, Personality and Emotions in User-Adapted Interaction*, 2001.

[2] G. Mark and A. Voss, "Attractivity in virtual environments: getting personal with your agent," in *Proc. of AAAI fall symposium on socially intelligent agents*, 1997.

[3] L.G. Farkas, "Linear proportions in above-and below-average women's faces," in *Anthropometric facial proportions in medicine*, Charles C. Thomas, Ed., pp. 119–129. Springfield, 1987.

[4] Leslie G Farkas and John C Kolar, "Anthropometrics and art in the aesthetics of women's faces.," *Clinics in plastic surgery*, vol. 14, no. 4, pp. 599–616, 1987.

[5] IS Penton-Voak, DI Perrett, and JW Peirce, "Computer graphic studies of the role of facial similarity in judgements of attractiveness," *Current Psychology*, vol. 18, no. 1, pp. 104–117, 1999.

[6] Parham Aarabi, Dominic Hughes, Keyvan Mohajer, and Majid Emami, "The automatic measurement of facial beauty," in *Systems, Man, and Cybernetics, 2001 IEEE International Conference on*. IEEE, 2001, vol. 4, pp. 2644–2647.

[7] R. White, A. Eden, and M. Maire, "Automatic prediction of human attractiveness," Tech. Rep., UC Berkeley, 2004.

[8] Hatice Gunes and Massimo Piccardi, "Assessing facial beauty through proportion analysis by image processing and supervised learning," *International journal of human-computer studies*, vol. 64, no. 12, pp. 1184–1199, 2006.

[9] Yael Eisenthal, Gideon Dror, and Eytan Ruppin, "Facial attractiveness: Beauty and the machine," *Neural Computation*, vol. 18, no. 1, pp. 119–142, 2006.

[10] Amit Kagian, Gideon Dror, Tommer Leyvand, Daniel Cohen-Or, and Eytan Ruppin, "A humanlike predictor of facial attractiveness," *Advances in Neural Information Processing Systems*, vol. 19, pp. 649, 2007.

[11] Jacob Whitehill and Javier R Movellan, "Personalized facial attractiveness prediction," in *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*. IEEE, 2008, pp. 1–7.

[12] Fangmei Chen and David Zhang, "A benchmark for geometric facial beauty study," in *Medical biometrics*. 2010, pp. 21–32, Springer.

[13] Davor Sutic, Ivan Breskovic, Rene Huic, and Ivan Jukic, "Automatic evaluation of facial attractiveness," in *MIPRO, 2010 Proceedings of the 33rd International Convention*. IEEE, 2010, pp. 1339–1342.

[14] Chih-Chung Chang and Chih-Jen Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, pp. 27, 2011.

[15] Douglas Gray, Kai Yu, Wei Xu, and Yihong Gong, "Predicting facial beauty without landmarks," in *Proceedings of the 11th European Conference on Computer Vision: Part VI*, Berlin, Heidelberg, 2010, ECCV'10, pp. 434–447, Springer-Verlag.

[16] Xuehan Xiong and F. De la Torre, "Supervised descent method and its applications to face alignment," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, June 2013, pp. 532–539.

[17] Jintu Fan, KP Chau, Xianfu Wan, Lili Zhai, and Ethan Lau, "Prediction of facial attractiveness from facial proportions," *Pattern Recognition*, vol. 45, no. 6, pp. 2326–2334, 2012.

[18] Michel F. Valstar, Hatice Gunes, and Maja Pantic, "How to distinguish posed from spontaneous smiles using geometric features," in *Proceedings of the 9th International Conference on Multimodal Interfaces*, New York, NY, USA, 2007, ICMI '07, pp. 38–45, ACM.

[19] G. Mckeown, M. F. Valstar, R. Cowie, M. Pantic, and M. Schroeder, "The semaine database: Annotated multimodal records of emotionally coloured conversations between a person and a limited agent," *IEEE Transactions on Affective Computing*, vol. 3, pp. 5–17, April 2012, Issue 1.

[20] Hatice Gunes, "A survey of perception and computation of human beauty," in *Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding*. ACM, 2011, pp. 19–24.

[21] Amir Navot, Lavi Shpigelman, Naftali Tishby, and Eilon Vaadia, "Nearest neighbor based feature selection for regression and its application to neural activity," in *Advances in Neural Information Processing Systems 18*. 2006, pp. 995–1002, MIT Press.