Personality Traits and Job Candidate Screening via Analyzing Facial Videos

Salah Eddine Bekhouche
Department of Electrical Engineering
University of Biskra, Algeria
salah@bekhouche.com

Fadi Dornaika
University of the Basque Country UPV/EHU, Spain
IKERBASQUE, Basque Foundation for Science, Spain
fadi.dornaika@ehu.es

Abdelkrim Ouafi
Laboratory of LESIA
University of Biskra, Algeria
a.ouafi@univ-biskra.dz

Abdelmalik Taleb-Ahmed
Laboratory of LAMIH
University of Valenciennes, France
taleb@univ-valenciennes.fr

Abstract

In this paper, we propose a novel approach for estimating the Big Five personality traits and the job candidate screening attribute through facial videos. At running time, the proposed system feeds the Pyramid Multi-Level (PML) texture features extracted from the whole video sequence to 5 Support Vector Regressors in order to estimate the personality traits. These estimated five scores are then considered as new input features to the interview score regressor. The latter is given by a Gaussian Process Regression (GPR). The experimental results on ChaLearn LAP APA2016 dataset achieve good performance. Furthermore, they demonstrate that the computational cost of both the training and the testing of the proposed framework are very competitive in terms of accuracy and computational cost.

1. Introduction

Computer vision technologies have been used in a lot of domains, one of the latest problem addressed in computer vision is to automatically evaluate the big five apparent personality traits (i.e. openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism) from videos of subjects speaking in front of a camera. This problem was addressed as a challenge called "ChaLearn Looking at People 2016 First Impressions challenge" [8], the term First Impressions refers to the first idea we get once we encounter new person. Due to the success of the last challenge a new challenge has been introduced which deals with a new problem about the job candidate screening attribute in addition to the Big Five personality traits. As a new research topic, approaches dealing with this problem are very few. In the sequel, we will describe briefly the works of the teams who participated in the last two challenges.

Zhang et al. [12] propose a framework called Deep Bimodal Regression (DBR) for the apparent personality analysis, the input video is considered as it has visual and audio modalities. Each of these two modalities is used to gives scores, the scores are fused in order to obtain final prediction of the personality traits.

Subramaniam et al. [9] propose two bi-modal deep neural network architectures for describing human personality which have two branches, the first one for encoding audio features and the other for visual features. They train these networks using temporally ordered audio and novel stochastic visual features from few frames.

Güçlütürk et al. [3] developed an audiovisual deep residual network for multi-modal apparent personality trait recognition. They train a convolutional network to predict the Big Five personality traits of people from, this network does not require face preprocessing.

Gürpınar et al. [4] proposed an approach to predict the first impressions that people will have when they encounter a new person using a short video. They used a pre-trained Deep Convolutional Neural Networks to extract features. Their approach relies on both the face and the whole video.

This paper addresses the estimation of the big five personality traits, also known as the five factor model (FFM) [11], and the interview score using facial videos. Unlike most of the previous works which were based on deep learning, our work is based on image textures. Although deep learning approaches can achieve better results, temporal face texre-based approaches are still very effective and not considered as time-consuming as deep learning.

The rest of the paper is organized as follows: in Section 2 we introduce our approach. The experiments and results are given in Section 3. In section 4 we give the conclusion and some future works.

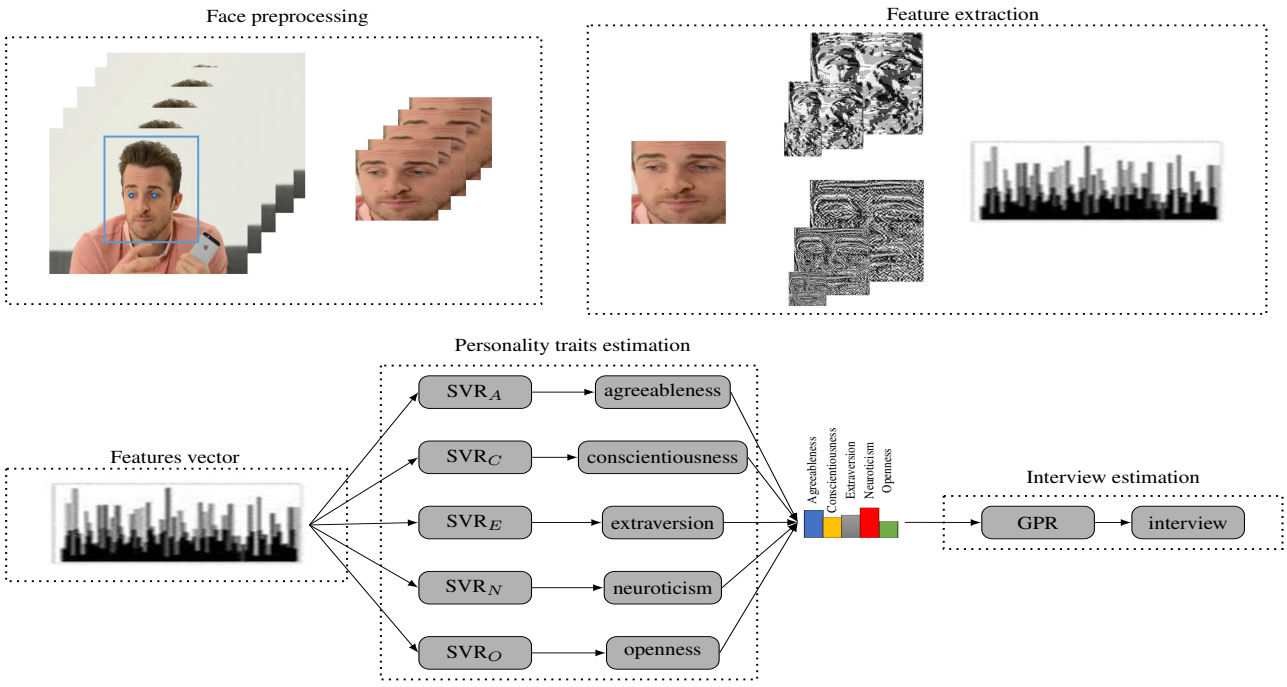


Figure 1. General structure of the proposed approach.

2. Methods

The proposed approach can be divided into three phases: face preprocessing, feature extraction and Personality traits and interview estimation. Figure 1 illustrates the proposed framework.

2.1. Face preprocessing

For each frame in the video, we apply the Haar cascade object detector that uses the Viola-Jones algorithm [10] in order to detect the face region. We then detect the face landmarks using Ensemble of Regression Trees (ERT) algorithm [6]. The locations of the two eyes are used to rectify the face 2D pose by applying a 2D similarity transform on the original face image [2]. Like in [1], we set the parameters $k_{side} = 0.5$, $k_{top} = 1$ and $k_{bottom} = 1.75$ to crop the face region of interest (ROI). Figure 2 illustrates the region of interest in the rectified face image.

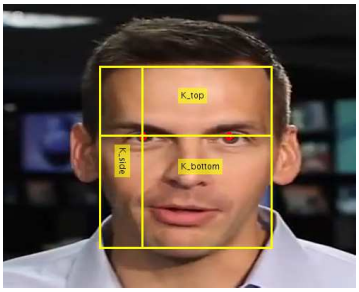


Figure 2. Cropping the face based on the eyes centers.

2.2. Feature extraction

The above 2D alignment is applied to each frame in the video sequence. After we obtain all the aligned faces, we apply two different texture descriptors, Local Phase Quantization (LPQ) [7] and Binarized Statistical Image Features (BSIF) [5], on each face, the results of the two descriptors will be represented by a Pyramid Multi-Level (PML). Figure 3 illustrates the PML principle for a pyramid of four levels when using the Local Phase Quantization (LPQ) as descriptor.

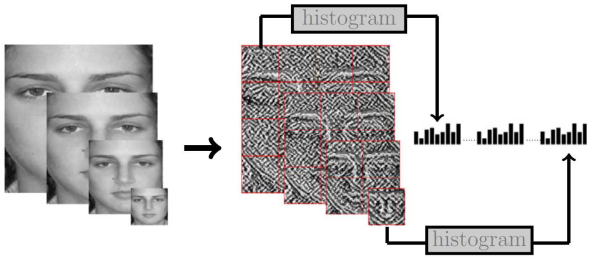


Figure 3. Pyramid Multi-Level Local Phase Quantization example of 4 levels.

The features extracted from both PML-BSIF and PML-LPQ are concatenated in one vector. Thus, each video frame has its own feature vector. In order to get the feature vector associated with the whole video sequence, we simply compute the mean of all feature vectors (see Figure 4). This allows to get a compact representation of the temporal information conveyed by the video sequence. In PML rep-

resentation, the original image is explicitly transformed and represented by several scaled images. Their number is equal to the number of levels. For each such an image, a specific grid is used to obtain a multi-block representation of the corresponding image [1].

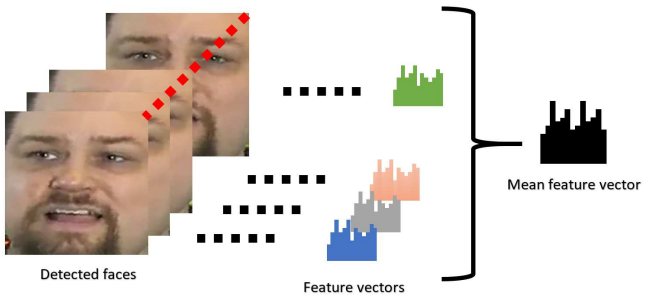


Figure 4. Representing the whole video sequence by a unified feature descriptor. This is carried out by computing the mean of feature vectors associated with the frames.

2.3. Personality traits and interview estimation

The proposed method first estimates the score for every main trait. These five estimated scores will be then used for estimating the interview score. Therefore, for scoring the personality traits we feed the PML features to five nonlinear Support Vector Regressors (SVRs) each SVR is for one of the big five personality traits.

These SVRs will separately estimate the 5 scores which are considered as new input features for the interview trait. We feed these features to a Gaussian Process Regression (GPR). Thus, we estimate interview score from the GPR based on the estimated five scores. We can note that the video sequence was not directly used by the interview score regressor. The GPR was tuned to find the best configuration for its hyper-parameters, namely the Sigma parameter and the kernel.

3. Experimental results

3.1. Data

In our experiments, we use the ChaLearn LAP 2016 APA database [8]. This database contains 10,000 videos which represent 41.6 hours (4.5M frames), this database was divided into three folds: a train fold having 60% of the videos, a validation fold having 20% of the videos, and a test fold having 20% of the videos.

In the first phase of the competition, the development phase, we had access to 6,000 labeled continuous video sequences (train) and 2,000 unlabeled continuous video sequences (validation). In the final phase, we had access to the labels of the validation videos and 2,000 unlabeled test videos.

3.2. Evaluation Protocol

For each video sequence, the ground truth for the big five traits and interview, are given by real scores that belong to the interval $[0, 1]$. The mean accuracy is used to evaluate the performance of each personality trait and the interview too, this term has been used in the previous works for the first impression challenge [8].

The mean accuracy for each trait is defined as:

$$A = 1 - \frac{1}{N}\sum_{i=1}^{N}|p_i - g_i| \quad (1)$$

where p_i is the predicted score, g_i is the ground truth score, and N is the total the total number of test videos.

3.3. Results and discussion

We conducted our experiments on ChaLearn LAP 2016 APA database. The approach was found very good for the trade-off efficiency-accuracy. Table 1 shows the CPU time of the proposed framework when applied on 6000 train videos and 2000 test videos. The experiments were carried out on a laptop DELL 7510 Precision (Xeon Processor E3-1535M v5, 8M Cache, 2.90 GHz, 64GB RAM, Ubuntu 16.04).

Table 1. CPU time of the proposed approach (Train/Test).

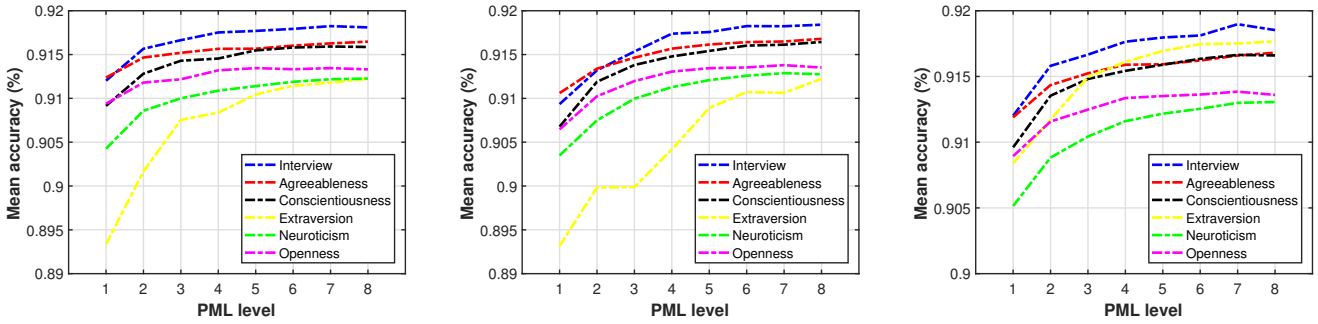
	Interview	Big Five traits	All
Train	13m	1h 36m	1h 49m
Test	1m	29m	30m
Total	14m	2h 5m	2h 19m

Figure 5 shows the results of personality traits and interview estimation obtained at the development phase when different PML levels and descriptors have been used. From these experiments, we observe that level 7 PML using mixed descriptors BSIF and LPQ gives the best results.

From this figure, we can observe that the performance associated with the interview score estimation is almost better than that associated with the big 5 personality traits despite the fact that the interview results are based on the other results.

The results of the quantitative stage of the challenge, the Job Candidate Screening Coopetition Challenge, are given in Table 2.

We were the second top-ranked team in both phases (Development and Final). Most of the participants used visual and sound features, in addition, the approaches were based on deep learning which needs more time for both training and testing. Our proposed framework seems to be more principled and its computational cost seems to be very reasonable.



a) BSIF b) LPQ c) BSIF+LPQ

Figure 5. Results of different descriptors using different PML levels.

Table 2. The results of the quantitative stage of the Job Candidate Screening Coopetition Challenge

Phase	Team	Interview	Agreeableness	Conscientiousness	Extraversion	Neuroticism	Openness
Development	BU-NKU	0.919782	0.916108	0.916633	0.920557	0.914871	0.916935
	PML (ours)	0.918974	0.916589	0.916639	0.917503	0.912992	0.913846
	FDMB	0.902961	0.908530	0.900877	0.904909	0.898365	0.903576
Final	BU-NKU	0.920916	0.913731	0.919769	0.921289	0.914613	0.917014
	PML (ours)	0.915746	0.910312	0.913775	0.91551	0.908297	0.910078
	ROCHCI	0.901859	0.903216	0.894914	0.90266	0.901147	0.904709

4. Conclusion

The paper presented a novel learning system for evaluating job candidates through facial videos. In this work, the big Five personality traits as well as the interview score are addressed. The image features were inspired from our recent work [1] on facial demographic estimation. The proposed approach uses these features to compute a temporal representation of these features from video sequence.

As future work, we envision the use of both audio and video features. We also envision to study and model the correlations between the big five personality traits and the decision of the interview.

References

[1] S. Bekhouche, A. Ouafi, F. Dornaika, A. Taleb-Ahmed, and A. Hadid. Pyramid multi-level features for facial demographic estimation. *Expert Systems with Applications*, 80:297 – 310, 2017.

[2] S. Bekhouche, A. Ouafi, A. Taleb-Ahmed, A. Hadid, and A. Benlamoudi. Facial age estimation using bsif and lbp. In *Proceeding of the first International Conference on Electrical Engineering ICEEB14*, 2014.

[3] Y. Güçlütürk, U. Güçlü, M. A. J. van Gerven, and R. van Lier. *Deep Impression: Audiovisual Deep Residual Networks for Multimodal Apparent Personality Trait Recognition*, pages 349–358. Springer International Publishing, Cham, 2016.

[4] F. Gürpınar, H. Kaya, and A. A. Salah. *Combining Deep Facial and Ambient Features for First Impression Estimation*, pages 372–385. Springer International Publishing, Cham, 2016.

[5] J. Kannala and E. Rahtu. Bsif: Binarized statistical image features. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pages 1363–1366, Nov 2012.

[6] V. Kazemi and J. Sullivan. One millisecond face alignment with an ensemble of regression trees. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1867–1874, June 2014.

[7] V. Ojansivu and J. Heikkilä. *Blur Insensitive Texture Classification Using Local Phase Quantization*, pages 236–243. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[8] V. Ponce-López, B. Chen, M. Oliu, C. Corneanu, A. Clapés, I. Guyon, X. Baró, H. J. Escalante, and S. Escalera. *ChaLearn LAP 2016: First Round Challenge on First Impressions - Dataset and Results*, pages 400–418. Springer International Publishing, Cham, 2016.

[9] A. Subramaniam, V. Patel, A. Mishra, P. Balasubramanian, and A. Mittal. *Bi-modal First Impressions Recognition Using Temporally Ordered Deep Audio and Stochastic Visual Features*, pages 337–348. Springer International Publishing, Cham, 2016.

[10] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–511–I–518 vol.1, 2001.

[11] Wikipedia. Big five personality traits — wikipedia, the free encyclopedia, 2017. [Online; accessed 8-April-2017].

[12] C.-L. Zhang, H. Zhang, X.-S. Wei, and J. Wu. *Deep Bimodal Regression for Apparent Personality Analysis*, pages 311–324. Springer International Publishing, Cham, 2016.