Landmark-free smile intensity estimation

Júlio César Batista, Olga R. P. Bellon and Luciano Silva
IMAGO Research Group - Universidade Federal do Paraná
{julio.batista,olga,luciano}@ufpr.br

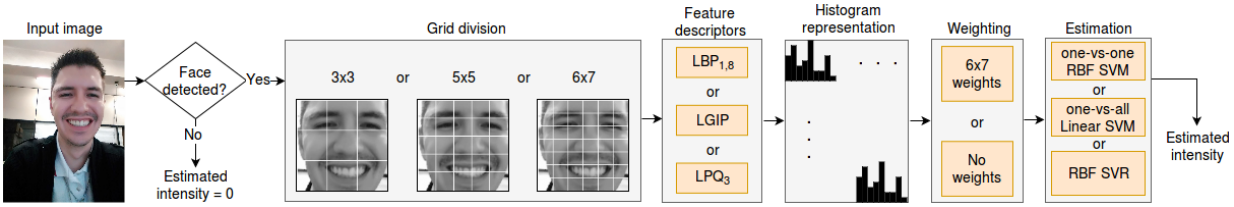

Fig. 1. Overview of our method for smile intensity estimation

***Abstract*—Facial expression analysis is an important field of research, mostly because of the rich information faces can provide. The majority of works published in the literature have focused on facial expression recognition and so far estimating facial expression intensities have not gathered same attention. The analysis of these intensities could improve face processing applications on distinct areas, such as computer assisted health care, human-computer interaction and biometrics. Because the smile is the most common expression, studying its intensity is a first step towards estimating other expressions intensities. Most related works are based on facial landmarks, sometimes combined with appearance features around these points, to estimate smile intensities. Relying on landmarks can lead to wrong estimations due to errors in the registration step. In this work we investigate a landmark-free approach for smile intensity estimation using appearance features from a grid division of the face. We tested our approach on two different databases, one with spontaneous expressions (BP4D) and the other with posed expressions (BU-3DFE); results are compared to state-of-the-art works in the field. Our method shows competitive results even using only appearance features on spontaneous facial expression intensities, but we found that there is still need for further investigation on posed expressions.**

***Keywords*-Smile intensity estimation; Facial expression analysis; Feature extraction; Machine learning;**

I. Introduction

Facial expressions play an important role on communication and can provide clues about emotions and intentions [1], [2]. Machine understanding of facial expressions would open new possibilities in areas such as human-computer interaction, affective computing, and health care [3], [4], [5], [6]. However, there are still many challenges to be overcome.

Current research is beginning to focus on what can be learned from facial expression's intensities [1], specially from smiles. Abel and Kruger [7] found that there is a relationship between the smile intensity in photographs and longevity. Kraus and Chen [8] predicted that prior to a physical confrontation, smiles' intensity might indicate reduced intentions to engage in hostile and aggressive actions. Stratou et al. [6] analyzed males and females subjects and found that males afflicted by post-traumatic stress disorder have shown less intense smiles than the female subjects. They also suggest that intensity of expressions such as anger, disgust, contempt and joy might be a measure of affect.

Although researches in facial expression analysis have been making a lot of progress towards facial expression recognition, it is unclear how to estimate the intensity of these facial expressions [1]. The analysis of facial expressions intensity might lead to better human-computer interfaces and help computer assisted health care [4], [6]. Smiles are, perhaps, the most studied [8] and the most common facial expression, specially in single images such as photographs [9]. Because of this, the estimation of smile intensity is a first step for future research towards facial expression intensity estimation.

Facial expression recognition is mostly based on Paul Ekman's Facial Action Coding System (FACS) [10] and its components, the Action Units (AUs). AUs correspond to the contraction of specific facial muscles and may vary in intensity [1]. The intensity ranges in a 5-level scale from 1 to 5 (or A to E), being 1 (A) the minimum intensity and 5 (E) the maximum intensity [1], [4]. There is also a sixth intensity level, which is 0, that indicates no AU activation [1], [4]. Fig. 2 shows an example of a 6-level smile intensity. These levels of intensity are also important in analyzing facial expression dynamics (changes of intensity over time) which goal is to detect onset, apex and offset of facial expressions [1], [2]. Detection of AUs can be handled using geometric features from facial fiducial points (landmarks) [11], appearance features [2], [12], or a combination of both [1], [13].

Girard et al. [1] evaluated the use of appearance features for smile intensity estimation. They extracted Gabor wavelets and SIFT [14] descriptors in regions surrounding the landmarks of the whole face (around 60). After feature extraction, they used Laplacian Eigenmap and Principal Component Analysis (PCA) for dimensionality reduction. For the prediction of smile intensity they evaluated Support Vector Regression (SVR), binary Support Vector Machines (SVMs), and multi-

Fig. 2. Examples of intensities from 0 (upper-left) to 5 (bottom-right)

class SVMs; all of them used a Radial Basis Function (RBF) kernel. They evaluated their method with spontaneous expressions from the BP4D [15] and the Spectrum [16] databases.

Nicolle et al. [13] worked on facial action unit prediction using shape and appearance features. The shape features were extracted from 49 landmarks. The appearance features were extracted dividing the image in a 4 x 4 grid and from 10 regions centered on landmarks to capture expression-related wrinkles. The classification of the intensity was handled using Lasso-regularization of Metric Learning for Kernel Regression where they aimed at reducing overfitting issues. Their method was evaluated using spontaneous expressions databases [17].

Jiang et al. [2] used the Local Phase Quantization from Three Orthogonal Planes (LPQ-TOP) descriptor for analysis of facial expressions dynamics. They detected a set of reference points to align the face to eliminate in-plane head rotation and address individual differences in face shapes. After the pre-processing, LPQ-TOP features are extracted from small blocks of the face. The detection of AUs was handled using SVM and GentleBoost for feature selection. Their method was evaluated in various databases, e.g., the Cohn-Kanade [18] and the UNBC-McMaster pain database [19], among others.

Contributions: We investigate the estimation of both posed and spontaneous smiles intensity without using landmarks. We evaluate a set appearance descriptors combined with machine learning models to describe these intensities. Our results favorably compared to state-of-the-art approach.

The rest of this paper is organized as follows: Section II describes our approach. Section III discusses experiments for both spontaneous and posed smile intensity estimation, followed by our conclusions in Section IV.

II. Smile intensity estimation

Face landmarks are prone to tracking failure [13], because of this we evaluate a set of appearance descriptors for smile intensity estimation. In this section we describe our approach in three steps: preprocessing, feature extraction and estimation. An overview of our method is shown in Fig. 1.

A. Preprocessing

The input to our method is a single image containing or not a smile. Firstly, we convert the image to grayscale and, to boost the face detection [20], resize the image to a width of 256 pixels with the height calculated keeping the aspect ratio in accordance to the width. If the face is detected, we crop the face region and scale it to 128x128 pixels [1].

TABLE I
Machine learning models' parameters

Parameter	Values
SVM C	2^0 to 2^{10}
RBF SVM γ	$\frac{1}{n_features}$, 2^0 to 2^4
SVR ϵ	0.1, 0.2, 0.4

B. Feature extraction

Our method doesn't rely on landmarks, so we look for appearance-features to describe smile intensity. Several appearance-based descriptors have been successfully used in facial expression recognition, some of them are Local Binary Pattern (LBP) [21], Local Gradient Increasing Pattern (LGIP) [22], and Local Phase Quantization (LPQ) [23]. A common approach when dealing with these descriptors is to segment the face in a grid shape. The grid size might vary from 3x3 to 10x10 [2], [4], [9], [12], [13]. Other approaches suggest the use of a weighted grid, or boosting, to give more importance to some regions than others [24], [25].

We evaluated the use of $LBP_{1,8}$, LPQ N_3, and LGIP on different grid sizes: 3x3 and 5x5 without weighting any region and 6x7 grid with weights as in [25]. For each cell we applied the appearance descriptors independently and extracted a normalized 256 bin histogram from the descriptor resulting image. In the case of 6x7 grid, each histogram is multiplied by the corresponding weight. Finally, the histogram of each cell is concatenated into a feature vector.

C. Estimation

The last step in our approach is the estimation of the facial expression intensity. This is a multiclass classification problem, because the intensity labels are discrete values and range from 0 to 5. SVMs are binary classifiers, but can be extended to multiclass classification using one-vs-all or one-vs-one decision schemes [26]. The classification models we selected were the one-vs-one RBF SVM as in [1], and one-vs-all Linear SVM. The one-vs-all was selected because it builds a model for each label, which is trained with instances of one label as positive samples against instances of all other labels as negative samples [26]. We also evaluated Support Vector Regression (SVR) as in [1], as for a classification task. The output of SVR is a continuous value from 0 to 5, converted to discrete values by rounding to the nearest integer.

III. Experiments and results

In this section we describe our experiments and the databases we selected. We discuss the results from our experiments and the limitations of our work.

A. Spontaneous smile intensity estimation

In this experiment we evaluated the performance of our approach for spontaneous smile intensity estimation using the BP4D database [15].The Spectrum Database also used by

TABLE II
ACCURACY SCORES OF OUR APPROACH FOR SPONTANEOUS SMILES

Model / Descriptor	LBP (5x5)	LGIP (3x3)	LPQ (3x3)
one-vs-all Linear SVM	605 (45%)	805 (60%)	920 (68%)
one-vs-one RBF SVM	778 (58%)	1088 (81%)	1104 (82%)
RBF SVR	692 (51%)	934 (69%)	986 (73%)

TABLE III
CONFUSION MATRIX FOR SPONTANEOUS SMILE INTENSITY ESTIMATION USING LPQ AND ONE-VS-ONE RBF SVM

Intensities	0	1	2	3	4	5
0	200	14	9	1	1	0
1	7	190	25	3	0	0
2	2	28	171	23	1	0
3	0	1	31	162	30	1
4	0	0	2	31	172	20
5	0	1	0	0	15	209

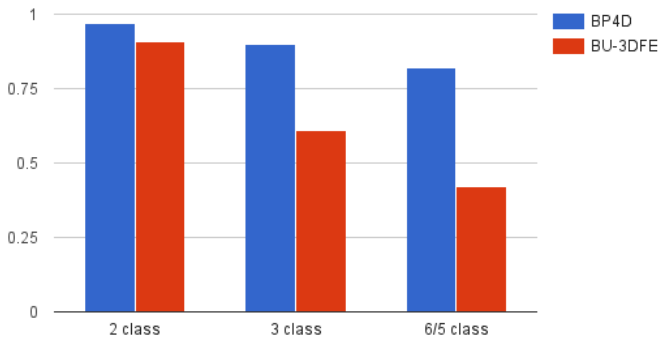

Fig. 3. Comparison of the 3x3 LPQ descriptor between spontaneous and posed smiles

[1] is not publicly available. The BP4D database contains recordings of 41 subjects engaging in 8 tasks to demonstrate spontaneous facial expressions. The subjects are distributed in about 56% female and 44% male. The database also provides FACS coding for each frame and intensity ground truth for some AUs, such as AU 12 and AU 14.

For this experiment we selected a subset of 9,000 images of the BP4D database, those presenting intensity levels for AU 12 (lip corner puller) which characterize a smile [5]. The subset contains a uniform distribution of intensities, for each intensity (0 to 5) 1,500 images were selected. These images were also uniformly split in three sets: training (70%), validation (15%), and testing (15%). A grid-search procedure was executed using the training set and the validation set with the parameters shown in Table I. After the selection of the parameters, we ran the experiments on the test set.

The general accuracy for this experiment is shown in Table II, presenting the accuracy on the test set using the descriptors and models that leaded to the best results using the validation set. From this table we can notice that our best result was the LPQ descriptor using a one-vs-one RBF SVM with a grid of 3x3 cells. We can also see that different descriptors yielded their best results using different grid sizes, such as 3x3 on LGIP and 5x5 on LBP. These results are also in agreement with the ones from [1], as the one-vs-one RBF SVM provided better results than the RBF SVR.

We calculated the Intraclass Correlation Coefficient (ICC) [27], using a 3x3 LPQ with one-vs-one RBF SVM, as a measure of agreement between our predictions and the ground truth. Our ICC(3, 1) score is 0.95 and is comparable to the scores in [1]. From the confusion matrix shown in Table III we can see the that the estimation errors were around the true label, in agreement with the ICC score.

B. Posed smile intensity estimation

In this experiment we evaluated the performance of our approach for estimation of posed smiles using the BU-3DFE database [28]. This database contains 3D models and 2D facial textures for facial expression analysis. The database contains 2,500 frontal face textures of 100 subjects. The subjects are distributed in about 60% female and 40% male. Each subject performed seven expressions (neutral, happiness, disgust, fear, angry, surprise and sadness). For each expression, in exception to neutral, there are four levels of intensity.

For this experiment we selected a subset of 1,000 images of the BU-3DFE database. These are all the images of the database that present the happy expression and other 200 images that do not show a smile. The images are uniformly distributed based on the intensities (0 to 4). We used the same setup as the previous experiment, we split the subset in 70% of the images for training, 15% for validation, and 15% for testing. As in Section III-A, a grid-search procedure was executed using the training set and the validation set with the parameters shown in Table I. After the selection of the parameters, we ran the experiments on the test set.

The general accuracy for this experiment is shown in Table IV. The results of this experiment are worst than the ones for spontaneous expressions. The best result in this experiment was 42% using a 3x3 LPQ with a one-vs-all Linear SVM classifier. This result is near the worst result for spontaneous smiles. These scores might be justifiable due to low amount of images used for training and testing. Using more images might increase the scores for posed smiles. We also calculated the ICC(3, 1) score for the best result in this experiment, the score is 0.67 which a low agreement between our predictions and the ground truth.

Given the results in the confusion matrix shown in Table III, we also tested for 2-class and 3-class groupings. The 2-class was executed as a measure of smile detection, with a label 0 meaning "no smile" and a label greater than 0 meaning "smile". For the 3-class we grouped the intensities as 0 and 1, 2 and 3, 4 and 5 to get better groupings. Fig. 3 shows the comparison of the LPQ descriptor between posed and spontaneous smiles. From this chart we can see that small groups yielded better results, suggesting that the classes in a 6-class grouping don't have well defined boundaries.

C. Limitations and future work

The main limitation of our approach is that we didn't investigate any alternative to handle head-pose variation. Recent methods, such as [29], could be used for face detection and,

TABLE IV
ACCURACY SCORES OF OUR APPROACH FOR POSED SMILES

Model / Descriptor	LBP (3x3)	LGIP (6x7)	LPQ (3x3)
one-vs-all Linear SVM	39 (26%)	61 (41%)	63 (42%)
one-vs-one RBF SVM	34 (23%)	45 (30%)	52 (35%)
RBF SVR	40 (27%)	60 (40%)	56 (37%)

before feature extraction, an alignment or face frontalization [30] step could be used. Another approach would be the use of face parts [31]. It is important to note that our best results were yielded by the combination of a grid division and appearance features without any weighting or boosting. We applied the weights suggested in [25], but our faces weren't aligned which might be the cause of the better performance without weighting. Another approach would be the selection of weights of our training set for a better selection of the most important regions. Although our approach performed well on spontaneous smiles, when applied to posed smiles the performance dropped to an accuracy of less than 50%. It indicates the need for better feature description and investigation on posed smiles, and also that our approach might be useful for detection of posed and spontaneous smiles. A downside of low-level features is that they might be affected negatively by identity bias [31] and a cross-database experiment would lead to more accurate results.

IV. FINAL REMARKS

In this paper we investigated the use of appearance features, techniques to enhance these features, and machine learning models for both posed and spontaneous smile intensity estimation. Even using only appearance features, our approach was able to estimate smile intensities with high accuracy. Our results are competitive with the state-of-the-art works for spontaneous smiles intensity, even relying only on appearance features. It is important to note that our findings suggest that the use of a simple grid division of the face led to better results than the ones using a weighted grid division. Our results show that boundaries in posed smiles are not well defined as in spontaneous ones because of the drop in performance in different groupings. Finally, our findings suggest that there is still need for improvement on estimation of posed smiles.

ACKNOWLEDGMENT

The authors would like to thank CAPES and CNPq for supporting this research.

REFERENCES

[1] J. M. Girard, J. F. Cohn, and F. D. la Torre, "Estimating smile intensity: A better way," *Pattern Recognition Letters*, vol. 66, pp. 13 – 21, 2015.
[2] B. Jiang, M. Valstar, B. Martinez, and M. Pantic, "A dynamic appearance descriptor approach to facial actions temporal modeling," *IEEE Transactions on Cybernetics*, vol. 44, no. 2, pp. 161–174, 2014.
[3] M. Pantic, "Machine analysis of facial behaviour: Naturalistic and dynamic behaviour," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, pp. 3505–3513, 2009.
[4] S. Kaltwang, O. Rudovic, and M. Pantic, *Continuous Pain Intensity Estimation from Facial Expressions*. Springer Berlin Heidelberg, 2012, pp. 368–377.
[5] S. Du, Y. Tao, and A. M. Martinez, "Compound facial expressions of emotion," *Proc. National Academy of Sciences*, vol. 111, no. 15, pp. E1454–E1462, 2014.
[6] G. Stratou, S. Scherer, J. Gratch, and L.-P. Morency, "Automatic nonverbal behavior indicators of depression and ptsd: the effect of gender," *Journal on Multimodal User Interfaces*, vol. 9, no. 1, pp. 17–29, 2015.
[7] E. L. Abel and M. L. Kruger, "Smile intensity in photographs predicts longevity," *Psychological Science*, vol. 21, no. 4, pp. 542–544, 2010.
[8] M. W. Kraus and T.-W. D. Chen, "A winning smile? smile intensity, physical dominance, and fighter performance." *Emotion*, vol. 13, no. 2, p. 270, 2013.
[9] C. C. Queirolo, M. P. Segundo, O. R. P. Bellon, and L. Silva, "Noise versus facial expression on 3d face recognition," in *ICIAP*, 2007.
[10] P. Ekman, W. Friesen, and J. Hager, *Facial Action Coding System (FACS): Manual*. A Human Face, 2002.
[11] M. F. Valstar, I. Patras, and M. Pantic, "Facial action unit detection using probabilistic actively learned support vector machines on tracked facial point data," in *IEEE CVPR Workshops*, 2005, pp. 76–76.
[12] B. Jiang, M. F. Valstar, and M. Pantic, "Action unit detection using sparse appearance descriptors in space-time video volumes," in *IEEE FG*, 2011, pp. 314–321.
[13] J. Nicolle, K. Bailly, and M. Chetouani, "Real-time facial action unit intensity prediction with regularized metric learning," *Image and Vision Computing*, vol. 52, pp. 1 – 14, 2016.
[14] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.
[15] X. Zhang, L. Yin, J. F. Cohn, S. Canavan, M. Reale, A. Horowitz, P. Liu, and J. M. Girard, "Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database," *Image and Vision Computing*, vol. 32, no. 10, pp. 692 – 706, 2014.
[16] J. F. Cohn, T. S. Kruez, I. Matthews, Y. Yang, M. H. Nguyen, M. T. Padilla, F. Zhou, and F. D. la Torre, "Detecting depression from facial actions and vocal prosody," in *Int'l Conf. Affective Computing and Intelligent Interaction*, 2009, pp. 1–7.
[17] S. M. Mavadati, M. H. Mahoor, K. Bartlett, P. Trinh, and J. F. Cohn, "Disfa: A spontaneous facial action intensity database," *IEEE Transactions on Affective Computing*, vol. 4, no. 2, pp. 151–160, 2013.
[18] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *IEEE FG*, 2000, pp. 46–53.
[19] P. Lucey, J. F. Cohn, K. M. Prkachin, P. E. Solomon, and I. Matthews, "Painful data: The unbc-mcmaster shoulder pain expression archive database," in *IEEE FG*, 2011, pp. 57–64.
[20] D. E. King, "Dlib-ml: A machine learning toolkit," *Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009.
[21] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on PAMI*, vol. 24, no. 7, pp. 971–987, 2002.
[22] L. Zhou and H. Wang, "Local gradient increasing pattern for facial expression recognition," in *IEEE ICIP*, 2012, pp. 2601–2604.
[23] V. Ojansivu and J. Heikkilä, *Blur Insensitive Texture Classification Using Local Phase Quantization*. Springer Berlin Heidelberg, 2008, pp. 236–243.
[24] G. Zhang, X. Huang, S. Z. Li, Y. Wang, and X. Wu, "Boosting local binary pattern (lbp)-based face recognition," in *Proc. Chinese Conf. Advances in Biometric Person Authentication*, ser. SINOBIOMETRICS'04, 2004, pp. 179–186.
[25] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803 – 816, 2009.
[26] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.
[27] P. Shrout and J. Fleiss, "Intraclass correlations: Uses in assessing rater reliability," *Psychological Bulletin*, vol. 86, no. 2, pp. 420–428, 1979.
[28] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3d facial expression database for facial behavior research," in *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, 2006, pp. 211–216.
[29] S. Liao, A. K. Jain, and S. Z. Li, "A fast and accurate unconstrained face detector," *IEEE Transactions on PAMI*, vol. 38, no. 2, pp. 211–223, 2016.
[30] T. Hassner, S. Harel, E. Paz, and R. Enbar, "Effective face frontalization in unconstrained images," in *IEEE CVPR Workshops*, 2015, pp. 4295–4304.
[31] E. Sariyanidi, H. Gunes, and A. Cavallaro, "Automatic analysis of facial affect: A survey of registration, representation, and recognition," *IEEE Transactions on PAMI*, vol. 37, no. 6, pp. 1113–1133, 2015.