Image preprocessing techniques for facial expression classification

ANDRE LUIZ DA S. PEREIRA*, Prograf, (IC), Universidade Federal Fluminense (UFF), Brazil
LEANDRO A. F. FERNANDES, Prograf, (IC), Universidade Federal Fluminense (UFF), Brazil
AURA CONCI, VisualLab, Institute of Computing (IC), Universidade Federal Fluminense (UFF), Brazil

Facial expression classification is an essential aspect of human interaction, and its proper classification is considered fundamental for the future inclusion of multimedia psychology into human-computer interfaces. For this purpose, this article aims to identify a preprocessing pipeline capable of reduce the accuracy variance for facial expression classification. Thereunto, Extended Cohn-Kanade Dataset, Support Vector Machine, and Bag of Features were used in six pipelines. Compared to a pipeline with minimal preprocessing, the best results presented an accuracy variance reduction of 65.63%.

CCS Concepts: • **Computing methodologies → Image processing**; **Classification and regression trees**; **Object recognition**.

Additional Key Words and Phrases: Facial expression, Image processing, Support Vector Machine, Feature computation, Bag of features.

ACM Reference Format:
Andre Luiz da S. Pereira, Leandro A. F. Fernandes, and Aura Conci. 2022. Image preprocessing techniques for facial expression classification. In *LIQUE 2022: Life Improvement in Quality by Ubiquitous Experiences Workshop, together with IMX 2022: ACM International Conference on Interactive Media Experiences. June 22-24, 2022, Aveiro, Portugal*. ACM, New York, NY, USA, 4 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Facial expression results from voluntary and non-voluntary face muscles movement denoting an emotional state [6, 9, 12, 15]. Therefore facial expressions recognition is significant in human communication [1, 5]. Ekman and Rosenberg [3] developed a taxonomic system capable of characterizing human facial expressions. Thereby, six universal expressions were discovered: happiness, sadness, fear, disgust, anger, and surprise. Moreover, such expressions can be found in any known culture, including primitive tribes that rarely interact with modern society [2].

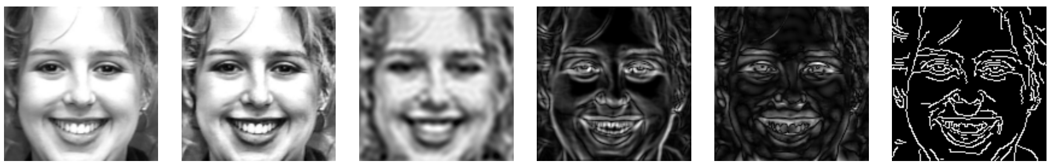
It is possible to say that the skin tonality, bone structure, face shape, and lighting are additional issues whose inclusion results in a considerable variety of the final facial appearance. For this reason, it is essential to preprocess facial images before the computation of features, discard unnecessary characteristics, and make the classification process as robust as possible. In such a direction, this work shows possible solutions to reduce the accuracy variation where there is a dataset with ethnic variety.

There are three more sections in this work: section 2 contains relevant points for classifying facial expressions; section 3 contains the experiments, results, and analyses; and section 4 shows the conclusions about the exploration.

2 CLASSIFICATION OF FACIAL EXPRESSIONS BY 2D TECHNIQUES

It is possible to perform the classification of facial expressions in different ways, and generally, it is executed with the following four steps.

The **input step** considers digital images included in the computer system (by acquisition or capture from cameras). There are several possible data sources for this step, such as: images of the upper face part [6]; 3D face laser scanning [7]; the stream of surveillance videos [1]; infrared, grayscale, or true-color images. However, we will consider common 2D images in the visible spectrum for this work. The database chosen for the experiments is the Extended Cohn-Kanade Dataset (CK+) [9]. CK+ is one of the most used databases for the classification of facial expressions, and its data follow the same standardization established by Ekman and Rosenberg [3].



(a) (b) (c) (d) (e) (f)

Fig. 1. Samples of the steps applied to CK+ dataset images. The images represent: (a) original face detected by Viola and Jones [16], cropping, resizing, and noise reduction with Gaussian filter; (b) equalization with CLAHE; (c) low pass filtering; (d) and (e) high pass filtering with Sobel and Fourier respectively; and (f) Canny edge detection.

The **preprocessing step** is used in almost all image processing applications to improve relevant characteristics for a specific problem. The most used preprocess approaches for facial expression classifications are: Noise Reduction; Face Detection; Normalization, and Histogram Equalization [5, 16]. Their usage aims to minimize disturbing details, find the face positioning and improve contrast. The preprocess applied in this work also follows this organization. In section 3 some tests are described to identify the best solution for our aim.

The **feature extraction step** computes the features or does characteristic extractions from the image. This extraction has two parts, definition of the number and type of the features and their computation from the preprocessed images. The first part is concerned with acknowledging the potential feature properties to see those that best represent the data to be analyzed. The final part presents their numerical results, which will be managed and used in diverse forms such as vectors, matrices, points in numerical or boolean space, and histograms. There are several techniques and combinations for both steps, such as, Local Binary Pattern (LBP) [4], Optical Flow [11], Bag of Features, method based on Scale Invariant Feature Transform (SIFT) [8], etc., each one with specific properties. In this work, the feature extraction was made by Bag of Features, that is a method that uses a set of training images to generate a set of local characteristics. When a new image is added, it extracts the characteristics by clustering it to find the set that most resembles the new feature set [13]. This way, only features that best represent the images in a generalized way are used in the subsequent classification process.

This work aims to identify a preprocessing pipeline that makes the information more concise. Therefore the **classification step** uses a method that does not belong to the family of deep learning ones. Common methods such as k-Nearest Neighbors (kNN) [17], Support Vector Machine (SVM) [14], Adaptive Boosting (Adaboost) [18], Bayesian [10], are examples of these. In this work, an SVM with a Gaussian kernel was employed for being a method with considerable resistance to overfitting, performs well with a reduced amount of data, and tends to converge to good results with few iterations.

Once the basic steps for classifying facial expressions have been presented, it is possible to discuss the experiments carried out and the methods used for them.

3 THE SIX PREPROCESSING PIPELINES AND THE EXPERIMENTS CARRIED OUT

Altogether, six preprocessing pipelines were analyzed. Each pipeline was computed in the same machine a hundred times to ensure a statistical relevance closer to the real one, and in each round random samples were chosen. From the dataset, 68 images per emotion were used, divided into 80% for training and the rest (20%) for validation. At the end of each round, the average validation accuracy was computed. To better understand the named steps in table 1, they are described below, detailing the techniques performed in the experiments. Note that the numbers that accompany them are merely for the organization. Their parameters were chosen by visually analyzing their results in experiments.

Step 1 considers the original image in 8 bits grayscale where the Viola and Jones [16] technique was applied to identify its region of interest (ROI). Then Gaussian filtering with a 3x3 kernel was used for noise reduction, and a resizing of 128x128 was applied (Figure 1(a)). **Step 2** is a histogram equalization with Contrast Limited Adaptive Histogram Equalization (CLAHE) using a threshold of two and considering an 8x8 pixel windowing (Figure 1(b)). **Step 3** is applied a Low Pass filtering in the frequency domain (after discrete Fourier transform), considering an elliptic area of 25% of the image's size (Figure 1(c)). In **Step 4**, Sobel filtering is applied using a 3x3 kernel and the same weight for vertical and horizontal (Figure 1(d)). **Step 5** High Pass filtering is performed in the frequency domain, considering an elliptical area of 10% of the image (Figure 1(e)). **Step 6** Canny edge detection is applied, considering 100 as the lower threshold and the upper threshold of 200 (Figure 1(f)).

Arranging the steps is possible to form the tested pipelines. Note that feature extraction and classification are the same for all of them. The High pass and Canny scenarios' performance is poor compared to the other scenarios; for this reason, both are discarded. Analyzing the average accuracy represented in Table 1, it is notable that four pipelines have very close performances (Original, CLAHE, Low Pass, and Sobel), between 70.8% and 69.3% accuracy. The named Original, where the images are almost intact, indicates that the other three (CLAHE, Low pass, and Sobel) where the images most suffered transformations did not impair the classification model. The robustness of these pipelines can be verified by the variance presented in Table 1. The best accuracy is achieved in the Original pipeline (70.8%), but it has the worst accuracy variance among the best four pipelines. Considering the accuracy and accuracy variance, Sobel is the best pipeline, containing 70.7% of accuracy and an accuracy variance of 3.01×10^{-2}. Sobel's accuracy is 0.1% lower (compared to Original), but its accuracy variance is 65.63% lower, justifying the tradeoff.

Table 1. Summary of the averaged results in each pipeline.

Pipelines	Definition	Accuracy in percentage %	Variance x10^2
Original	Step 1	70.80	8.76
CLAHE	Step 1 + Step 2	69.30	8.41
Low pass	Step 1 + Step 2 + Step 3	69.80	7.96
Sobel	Step 1 + Step 2 + Step 4	70.70	3.01
High pass	Step 1 + Step 2 + Step 5	58.20	13.16
Canny	Step 1 + Step 2 + Step 6	55.80	6.96

4 CONCLUSION

Based on Table 1, the Sobel pipeline presents the best results, it highlights face features such as contours, lines, and volume relevant to facial expressions classification and discard undesired ones, such as light changes and color. Making the data less susceptible to variations does not necessarily imply that the classification model achieves higher accuracy, but the overall result could present a smaller variation margin.

For future works, it is important to explore datasets with a more ethnic diversity to better compare the alternatives. Varying the used parameters is another work that can be done to analyze if it is possible to achieve a higher accuracy or a lower accuracy variance.

ACKNOWLEDGMENTS

Acknowledgments of the authors are for CAPES. AC thanks project CAPES/PRINT, CNPq Brazilian agencies Grants (305416 / 2018-9) and FAPERJ (projects SIADE-2, e-Health Rio, SNE and Digit3D).

REFERENCES

[1] Keith Anderson and Peter W McOwan. 2006. A real-time automated system for the recognition of human facial expressions. *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society* 36, 1 (2006), 96–105. https://doi.org/10.1109/TSMCB.2005.854502

[2] Paul Ekman and Wallace V Friesen. 1971. Constants across cultures in the face and emotion. *Journal of personality and social psychology* 17, 2 (1971), 124.

[3] Paul Ekman and Erika L Rosenberg. 1997. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS).* Oxford University Press, USA.

[4] Xiaoyi Feng, Matti Pietikäinen, and Abdenour Hadid. 2007. Facial expression recognition based on local binary patterns. *Pattern Recognition and Image Analysis* 17, 4 (2007), 592–598.

[5] Yunxin Huang, Fei Chen, Shaohe Lv, and Xiaodong Wang. 2019. Facial expression recognition: A survey. *Symmetry* 11, 10 (2019), 1189.

[6] James J. Lien, Jeffrey F. Cohn, Takeo Kanade, and Ching Chung Li. 1998. Automated facial expression recognition based on FACS action units. *Proceedings - 3rd IEEE International Conference on Automatic Face and Gesture Recognition, FG 1998* (1998), 390–395. https://doi.org/10.1109/AFGR.1998.670980

[7] Shuming Liu, Xiaopeng Chen, Di Fan, Xu Chen, Fei Meng, and Qiang Huang. 2016. 3D smiling facial expression recognition based on SVM. In *2016 IEEE International Conference on Mechatronics and Automation.* IEEE, 1661–1666.

[8] Yu Liu, Jing-dong Wang, and Peng Li. 2011. A feature point tracking method based on the combination of SIFT algorithm and KLT matching algorithm. *Journal of Astronautics* 32, 7 (2011), 1618–1625.

[9] Patrick Lucey, Jeffrey F Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. 2010. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on.* IEEE, 94–101.

[10] Qirong Mao, Qiyu Rao, Yongbin Yu, and Ming Dong. 2016. Hierarchical Bayesian theme models for multipose facial expression recognition. *IEEE Transactions on Multimedia* 19, 4 (2016), 861–873.

[11] Kenji Mase. 1991. Recognition of facial expression from optical flow. *IEICE transactions (E)* 74 (1991), 3474–3483.

[12] Albert Mehrabian and James A Russell. 1974. *An approach to environmental psychology.* the MIT Press.

[13] Stephen O'Hara and Bruce A Draper. 2011. Introduction to the bag of features paradigm for image classification and retrieval. *arXiv preprint arXiv:1101.3354* (2011).

[14] Jamal Hussain Shah, Muhammad Sharif, Mussarat Yasmin, and Steven Lawrence Fernandes. 2020. Facial expressions classification and false label reduction using LDA and threefold SVM. *Pattern Recognition Letters* 139 (2020), 166–173.

[15] Silvan S Tomkins. 1980. Affect as amplification: Some modifications in theory. *Emotion: Theory, research, and experience* 1 (1980), 141–164.

[16] Paul Viola and Michael Jones. 2001. 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*, Vol. 1. IEEE, I–I.

[17] Xiao-Hu Wang, An Liu, and Shi-Qing Zhang. 2015. New facial expression recognition based on FSVM and KNN. *Optik* 126, 21 (2015), 3132–3134.

[18] Song Zhang, Bin Hu, Tiantian Li, and Xiangwei Zheng. 2018. A study on emotion recognition based on hierarchical adaboost multi-class algorithm. In *International Conference on Algorithms and Architectures for Parallel Processing.* Springer, 105–113.