

Camera-based Real-time Emotion Classification Using Support-Vector Machines

Wout Swinkels¹, Luc Claesen¹, Shen Haibin²

¹Faculty of Engineering Technology, University Hasselt, Diepenbeek, Belgium

²College of Electrical Engineering, Zhejiang University, Hangzhou, China

Abstract— Extracting facial emotions is a great challenge in the area of computer vision, unlike humans, machines have a lot of difficulties with interpreting emotions. Due to the work of Viola-Jones [1] emotion detection is already more feasible. Using a fast algorithm to extract the region of interest (ROI) for emotion detection brings the development of a real-time emotion classification algorithm one step closer.

In this research an emotion detection algorithm aimed at real-time implementation on an embedded platform has been developed. Several algorithms have been implemented and tested to evaluate their results with respect to their real-time performance potential and classification efficiency. The developed algorithm relies on three fundamental findings, which relate to the way humans interpret emotions. These findings are:

- Face detection (region of interest detection).
- Feature selection and extraction, to use relevant features which distinguish different emotions
- Interpreting extracted features, classify emotions based on their features

The face detection in the developed algorithm is not based on the Viola-Jones algorithm but on the Histogram of Oriented Gradients (HoG) algorithm [2]. The reason is that the HoG algorithm outperforms the Viola-Jones algorithm once implemented on the platform designed for the emotion detection. Pre-processing the captured image before passing it to the HoG algorithm makes it possible to achieve face detection in less than 5 ms.

After face detecting, the features are extracted in the region of interest (ROI). First 19 points located around the eyebrows, eyes, nose and mouth are extracted from the ROI by using an ensemble of regression trees [3]. This is accomplished in less than 2 ms. Then the displacement ratio is calculated for 12 distances, as depicted in Figure 1, between these feature points. These displacement ratios are used as the features for the classification algorithm.



Figure 1: Image expressing surprise with the 12 distances (©Jeffrey Cohn).

For classifying the emotions multi-class support vector machines [4] are implemented in a cascade with SVMs for binary classification. This approach is chosen to improve the overall accuracy by training separate SVMs for emotions that are detected very weak. The results of this approach are significant with an average accuracy of 89,78% including outliers for the emotions of contempt and surprise. These results are obtained when deploying the proposed method on the Extended Cohn-Kanade Dataset (CK++) [5].

Comparing the accuracies of the developed algorithm with current state-of-the art algorithms [6] and [7] it is clear that the proposed algorithm has a lot of potential. This algorithm outperforms the other algorithms in the detection of 4 emotions with outliers for contempt and surprise, as indicated in Table 1. Another aspect is that this algorithm is a real-time algorithm because the total processing time is less than 30 ms.

Keywords—HoG; Ensemble of Regression Trees; Cascaded SVM classification; Emotion detection; Real-time performance

Table 1: Accuracy comparison.

	Proposed method	[6]	[7]
Anger	81,82%	71,39%	92,31%
Contempt	100%	/	/
Disgust	96,55%	95,33%	93,62%
Fear	91,67%	81,11%	90,57%
Happy	94,12%	95,42%	84,62%
Sadness	64,29%	88,01%	86,05%
Surprise	100%	98,27%	90,24%
Neutral	/	/	90,74%
Average	89,78%	88,26%	89,74%

[1] P. Viola and M. Jones, "Robust Real-time Object Detection," in *Second International Workshop on Statistical and Computational Theories of Vision - Modeling, Learning, Computing, and Sampling*, Vancouver, Canada, 2001.

[2] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 886-893, 2005.

[3] J. Sullivan and V. Kazemi, "One Millisecond Face Alignment with an Ensemble of Regression Trees," in *Conference on Computer Vision and Pattern Recognition*, Columbus, Ohio, 2014.

[4] V. N. Vapnik, *The Nature of Statistical Learning Theory*, New York: Springer-Verlag, 2000.

[5] T. Kanade, J. F. Cohn and Y. Tian, "Comprehensive database for facial expression analysis," *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00)*, pp. 46-53, 2000.

[6] L. Zhong, Q. Liu, P. Yang, B. Liu, J. Huang and D. N. Metaxas, "Learning Active Facial Patches for Expression Analysis," *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2562-2569, 2012.

[7] M. Song, D. Tao, Z. Liu, X. Li and M. Zhou, "Image Ratio Features for Facial Expression Recognition Application," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 3, pp. 779-788, 2010.