Spontaneous Emotion Recognition from Three-Dimensional Dynamic Facial Sequences

Jadisha Yarif Ramírez Cornejo, Helio Pedrini

Institute of Computing, University of Campinas, Brazil, 13083-852

Abstract

Facial expressions provide important indications about human emotions. The development of an automatic emotion recognition system based on facial expressions has received significant attention due to its applications in several knowledge domains, such as affective computing, behavior prediction, security, biometrics, and human-computer interactions. In this work, we propose a novel and automatic methodology for emotion recognition from dynamic facial expressions. Our approach encodes video appearance, shape, and motion information by combining Visual Rhythm (VR) and Motion History Image (MHI), further describing them through texture analysis. These texture analysis methods include Weber Local Descriptor (WLD), Census Transform Histogram (CENTRIST), Histogram of Oriented Gradients (HOG), and Gray-Level Co-Occurrence Matrix (GLCM). For assessing our methodology, we conducted experiments on the Binghamton University 4D Facial Expression (BU-4DFE) data set, leading to accuracy improvements on partial facial regions. In summary, we conclude that the combination of Visual Rhythm and Motion History Image aids on the automatic recognition of dynamic facial expressions.

Methodology

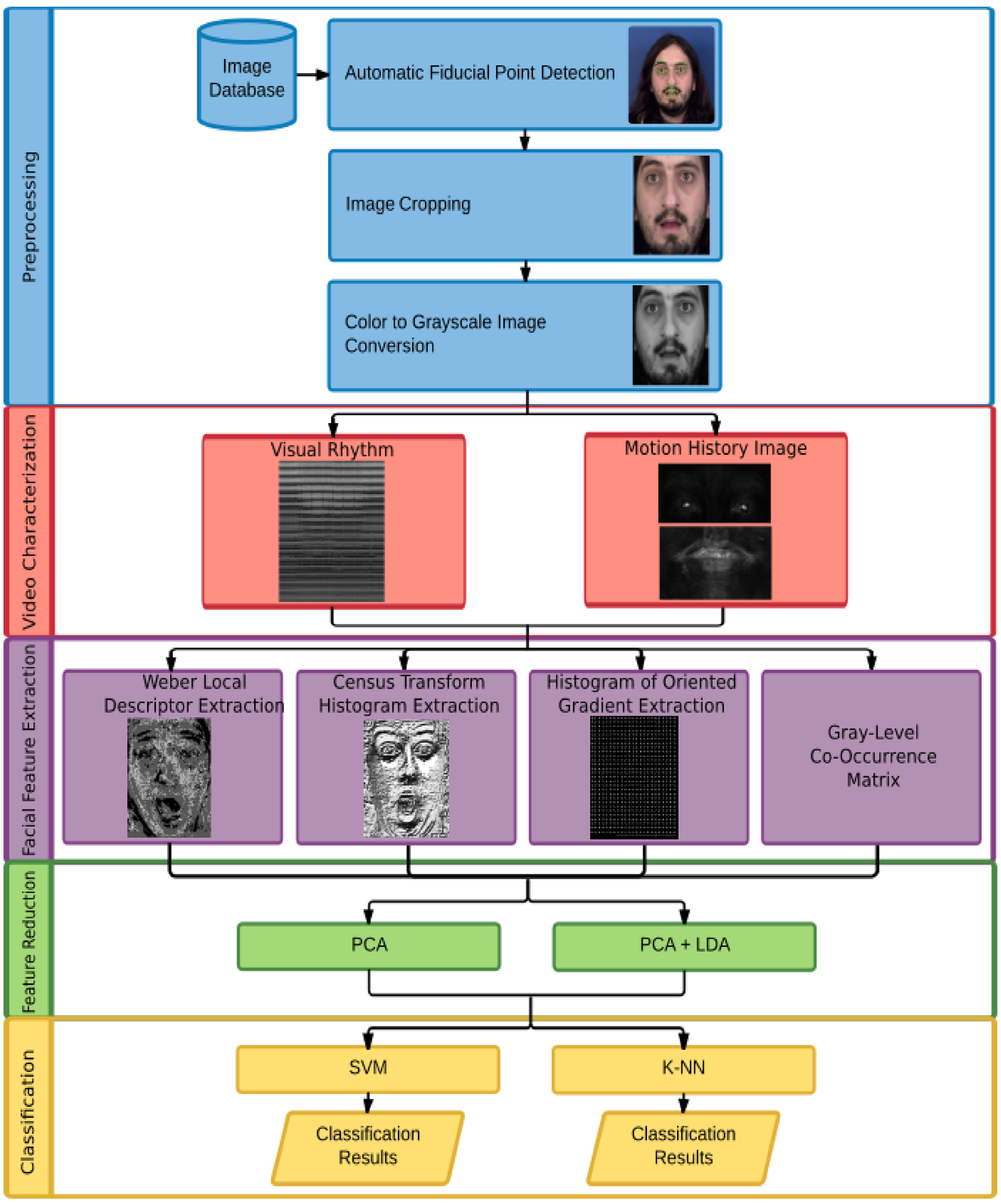
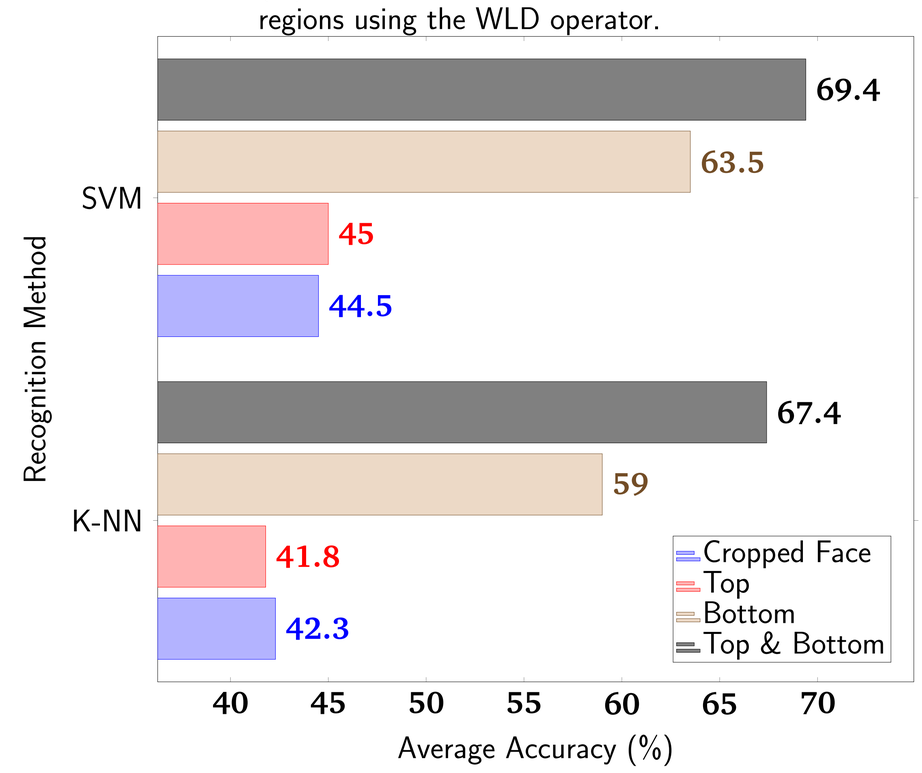


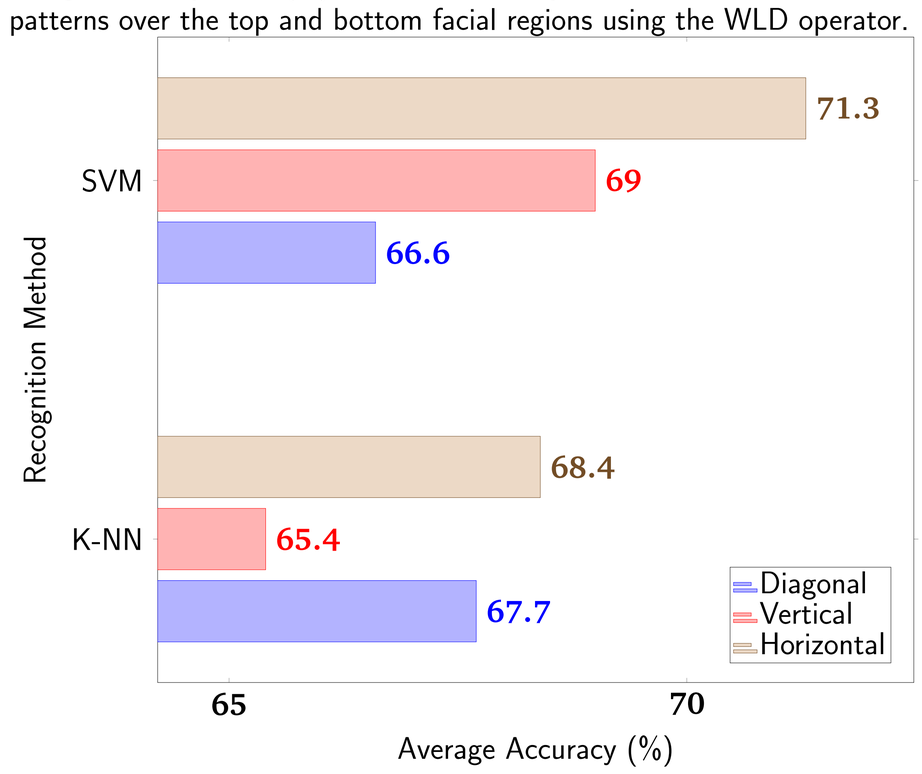
Diagram illustrating the dynamic emotion recognition methodology.

Experimental Results

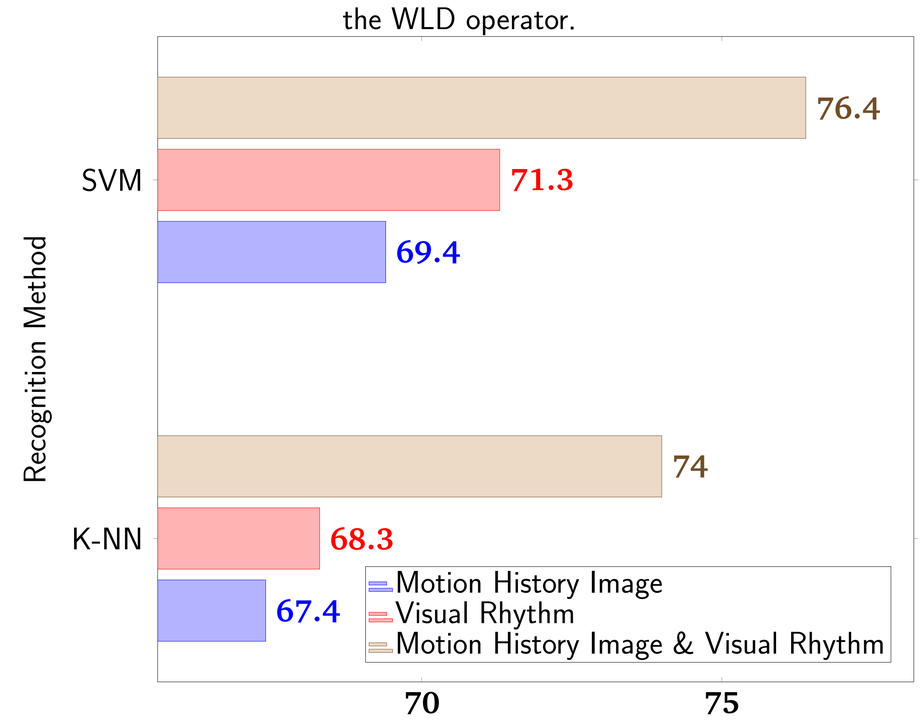
Average accuracy rates (%) applying Motion History Images over different facial regions using the WLD operator.



Average accuracy rates (%) for the Visual Rhythm with different orientation patterns over the top and bottom facial regions using the WLD operator.



Average accuracy rates (%) employing Motion History Images and Visual Rhythms (horizontal orientation) over the top and bottom facial regions, using the WLD operator.



Conclusions and Work in Progress

- Visual rhythm is introduced as a potential video characterization technique for dynamic emotion recognition.
- Experimental results have shown that fusing Visual Rhythms and Motion History Images influence positively in the recognition rates.
- The bottom facial expression region contains enough discriminative information for emotion recognition. Moreover, the combination of top and bottom regions presents even more representative information.
- We have conducted experiments based on Deep Learning techniques, that is, we have combined the feature vectors obtained with our approach and the features extracted from a 3D Convolutional Neural Network (3D CNN) followed by a Recurrent Neural Network (RNN).

jadisha.cornejo@ic.unicamp.br, helio@ic.unicamp.br