DEPTH VIDEO-BASED FACIAL EMOTION ANALYSIS

Daniel Barmaimon
Supervisors: Dr. Djamila Aouada*, Dr. David Fofi^

* Interdisciplinary Centre for Security, Reliability and Trust - University of Luxembourg
^ University of Bourgogne

ABSTRACT

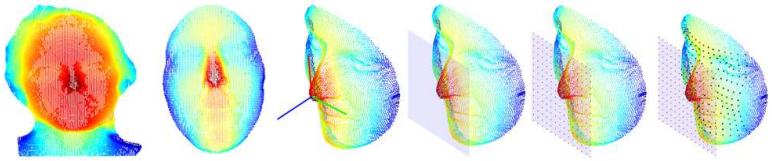
This paper presents a system to analyze spontaneous emotions from facial expressions captured with a 3D/depth sensing system. The possibility of controlling the resolution and noise level of the data allows to emulate off-the-shelf depth cameras. A whole pipeline is designed for the study of 3D depth video sequences of human faces. Three main steps define the pipeline: preprocessing, feature extraction and classification. In the preprocessing step the facial point cloud is aligned and reduced to a sorted group of 3D points, hence maintaining an invariant position with respect to the subject's face. This, in turn, allows a local temporal analysis. In the feature extraction step, the mean curvature for each of the points in the new cloud is calculated creating a 2D curvature map. The curvature map is divided into patches, over which a histogram analysis is performed. The bins of the histograms are the features of the system. The system is trained with classification models such as Support Vector Machines or binary trees, and validated with cross-validation. This system is designed as a tool for quantitative analysis of resolution enhancement and denoising algorithms, allowing to study the recognition rate before and after applying them in the context of facial expression recognition from 3D videos.

OBJECTIVES

- Creation of a system to detect facial emotions from 3D video sequences, based on local analysis of curvature.
- Ability to control the resolution and the level of noise on clean data, allowing to simulate the acquisition of different vision systems.
- Analysis over spontaneous emotions, that are natural and realistic, which could be used in real world applications.

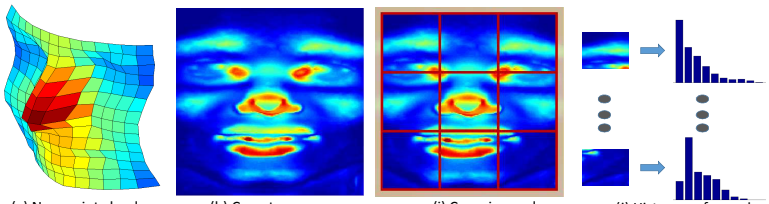
PIPELINE DESCRIPTION

PREPROCESSING



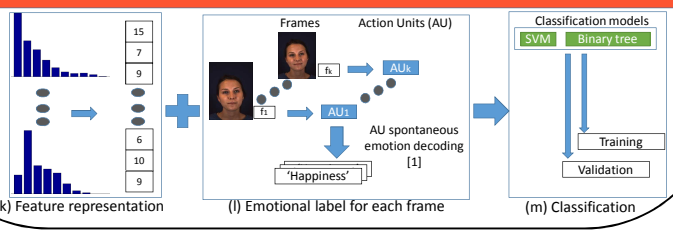
(a) Original point cloud (b) Cropped point cloud (c) Locating nose-tip & PCA (d) Plane definition (e) Grid generation (f) New point cloud interpolation

FEATURE EXTRACTION



(g) New point cloud (h) Curvature map (i) Cropping and division into patches (j) Histogram for each patch

CLASSIFICATION

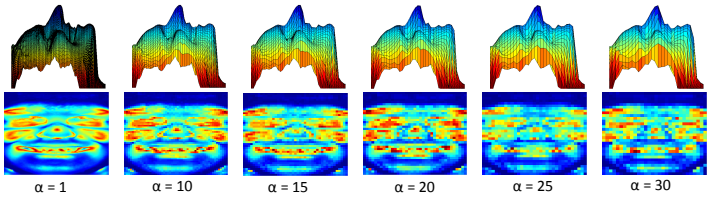


(k) Feature representation (l) Emotional label for each frame (m) Classification

RESULTS

1. Pipeline's analysis

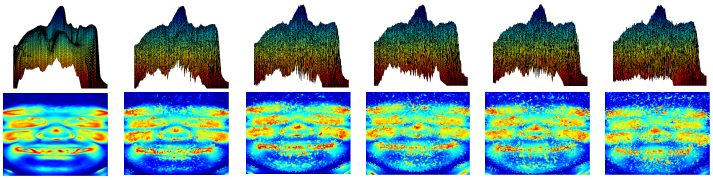
The resolution is controlled using $\alpha = \frac{N_{init}}{N_{new}}$ where N_{init} is the number of points of the cropped point cloud and N_{new} is the number of points in the interpolated point cloud.



α = 1 α = 10 α = 15 α = 20 α = 25 α = 30

Interpolated point cloud and curvature maps varying the α with noiseless data.

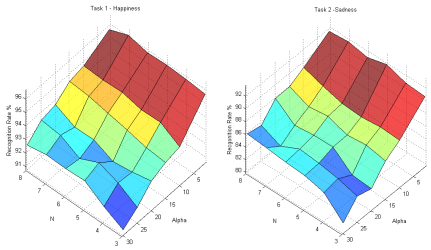
For evaluation of the results with noise a normal distribution is used in the three axis with mean μ = 0 and standard deviation σ.



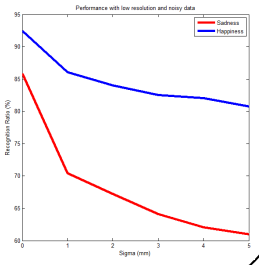
σ = 0 mm. σ = 1 mm. σ = 2 mm. σ = 3 mm. σ = 4 mm. σ = 5 mm.

Interpolated point cloud and the curvature maps varying the noise level and α = 1.

2. Performance varying the number of patches divisions



Curvature maps are divided into N x N patches. An analysis of the recognition rate for different values of N and α is performed.



3. Emulation of depth commodity camera

Resolution is reduced by α = 30 and the number of patches is optimized to N = 8, to emulate a PMD Camboard Nano time-of-flight camera [2]. Analysis of spontaneous emotions' recognition under these conditions.

CONCLUSIONS

- A whole system for analysis of spontaneous emotions using 3D/depth facial video sequences has been developed.
- The possibility to vary the resolution and the level of noise allows the emulation of current depth commodity cameras.
- Analysis on spontaneous emotions on facial expressions is complicated due the subtle changes that differentiate them. Some expressions, such as happiness, are essentially different by nature from any other.
- The current system can be used for the analysis of resolution enhancement and denoising algorithms, with the possibility of measuring differences quantitatively.

REFERENCES

[1] Xing Zhang, Lijun Yin, Jeffrey F Cohn, Shaun Canavan, Michael Reale, Andy Horowitz, Peng Liu, and Jeffrey M Girard. Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database. Image and Vision Computing, 32(10):692–706, 2014.

[2] PMD Technologies. http://www.pmdtec.com/products_services/reference_design.php.