

Depth Videos for the classification of Micro-Expressions

Ankith Jain Rakesh Kumar*, Bir Bhanu, Christopher Casey, Sierra Grace Cheung and Aaron Seitz University of California, Riverside

UNIVERSITY OF CALIFORNIA, RIVERSIDE

Introduction

- Facial Macro-expressions
 - \circ Prolonged
 - Easily Recognizable
 - Time Frame > 0.6 second
- > Facial Micro-Expressions
 - Subtle
 - Spontaneous
 - Time Frame \leq 0.6 seconds





Fear (Macro-Expression)

Fear (Micro-Expression)



Existing Datasets



Datasets	Subjects	Expressions	Videos	Resolution	fps	Mean Age Group	Activities	Data Format	Duration of Video (sec)	Comments
Polikovsky [1]	10	6	42	640x480	200	N/A	N/A	Grayscale	N/A	Posed Expressions, Very few videos, no FACS coding and No pose variations
USF-HD [2]	N/A	4	100	1280x720	30	N/A	N/A	RGB	0.66	Posed, only 4 expressions considered, no FACS coding and No pose variations
SMIC [3]	16	3	164	640x480	100	26.7	Movies, YouTube	RGB	0.5	Spontaneous, no FACS coding, and No pose variations
CASME II [4]	35	5	247	640x480	200	22	Movie, YouTube	RGB	0.5	Spontaneous, only 5 expressions and No pose variations
SAMM [5]	32	7	159	2040x1088	200	33.24	YouTube	Grayscale	0.5	Spontaneous and No pose variations
VISME	21	6	238	1280x720	30	22.3	Movie Clips	RGB +Depth	0.6	Spontaneous, RGB and Depth Videos included and Various poses of the face

[1] Polikovsky, S. et al. Facial micro-expressions recognition using high speed camera and 3d-gradients descriptor. In Conference on Imaging for Crime Detection and Prevention, 2009

[2] Shreve, M. et al, Macro-and micro-expression spotting in long videos using spatio-temporal strain," in Face and Gesture 2011, March 2011,

[3] Li, Xi. et. al, A Spontaneous Micro-expression Database: Inducement, collection and baseline, 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2013

[4] Yan, W. J. et. al. "Casme ii: An improved spontaneous microexpression database and the baseline evaluation," PLOS ONE, vol. 9, 2014

[5]. Davison, A. K. et. al, "Samm: A spontaneous micro-facial movement dataset," IEEE Transactions on Affective Computing, vol. 9, Jan 2018.

Why Depth Videos?



- \succ Depth videos are less prone to the illumination changes
- The advantage of using depth videos is that the pixel intensities in the depth frames are based on the distance of the face to the camera that provides new information about the facial features.
- Privacy concerns

VISME Dataset





Sad

Нарру

Surprise

Anger

Disgust

Technical Approach





UCR

Feature Extraction

- HOG3D
- CNN
- CNN-LSTM

Part-Based Approach

- Eye
- Nose
- Mouth
- > Full Face Approach
- 5 Folds Cross-Validation

Expressions	No. of Videos
Anger	60
Disgust	26
Нарру	55
Fear	23
Sad	22
Surprise	52
Total	238

Experimental Results: Depth Videos



Evaluation Metrics

- MA: Mean Accuracy (S.D.)
- MFAR: Mean False Alarm Rate (S.D.)

Method	Нарру		Surprise		Disgust		Anger		Fear		Sad		Overall	
Wiethou	MA	MFAR												
HOG3D	0.6211	0.2073	0.6366	0.2566	0.6978	0.1000	0.6650	0.2957	0.7342	0.2529	0.8714	0.2813	0.7044	0.2323
(Full Face)	(0.0211)	(0.0110)	(0.0210)	(0.0150)	(0.0241)	(0.0145)	(0.0147)	(0.0207)	(0.0402)	(0.0274)	(0.0119)	(0.0207)	(0.0223)	(0.0192)
HOG3D	0.6400	0.2026	0.6595	0.2270	0.6825	0.2653	0.6917	0.2731	0.8480	0.0984	0.8914	0.1347	0.7355	0.2002
(Part-Based)	(0.0232)	(0.0139)	(0.0113)	(0.0142)	(0.0194)	(0.0122)	(0.0184)	(0.0247)	(0.0310)	(0.0273)	(0.0290)	(0.0203)	(0.0221)	(0.0201)
CNN	0.6569	0.1219	0.6973	0.1812	0.6667	0.1745	0.7476	0.2236	0.9413	0.2898	0.7371	0.0780	0.7412	0.1782
(Full Face)	(0.0273)	(0.0107)	(0.0101)	(0.0180)	(0.0139)	(0.0137)	(0.0559)	(0.0198)	(0.0218)	(0.0520)	(0.0105)	(0.0400)	(0.0222)	(0.0239)
CNN	0.6442	0.1112	0.7000	0.1497	0.7021	0.2429	0.7389	0.2702	0.9461	0.2631	0.7400	0.0742	0.7452	0.1852
(Part-Based)	(0.0190)	(0.0138)	(0.0123)	(0.0103)	(0.0288)	(0.0193)	(0.0174)	(0.0167)	(0.0267)	(0.0283)	(0.0156)	(0.0437)	(0.0200)	(0.0220)
CNN-LSTM	0.6800	0.1069	0.7108	0.1170	0.7206	0.1359	0.7452	0.1342	0.8902	0.1287	0.8743	0.1347	0.7702	0.1262
(Full Face)	(0.0233)	(0.0193)	(0.0150)	(0.0143)	(0.0181)	(0.0174)	(0.0201)	(0.0208)	(0.0340)	(0.0297)	(0.0120)	(0.0412)	(0.0204)	(0.0238)
CNN-LSTM	0.6695	0.1009	0.7014	0.1082	0.7275	0.2020	0.7159	0.1217	0.8671	0.1287	0.8771	0.1382	0.7598	0.1333
(Part-Based)	(0.0147)	(0.0213)	(0.0088)	(0.0177)	(0.0235)	(0.0227)	(0.0124)	(0.0192)	(0.0457)	(0.0273)	(0.0128)	(0.0457)	(0.0196)	(0.0257)

Experimental Results : RGB and Depth Comparison

Evaluation Metrics

- MA: Mean Accuracy (S.D.)
- MFAR: Mean False Alarm Rate (S.D.)

Data Format	Нарру		Surprise		Disgust		Anger		Fear		Sad		Overall	
	MA	MFAR												
Depth	0.6695	0.1069	0.7014	0.1082	0.7275	0.2020	0.7159	0.1217	0.8671	0.1287	0.8771	0.1382	0.7598	0.1333
	(0.0147)	(0.0213)	(0.0088)	(0.0177)	(0.0235)	(0.0227)	(0.0124)	(0.0192)	(0.0457)	(0.0273)	(0.0128)	(0.0457)	(0.0196)	(0.0257)
RGB	0.6823	0.1273	0.7179	0.1224	0.7321	0.2189	0.7592	0.1472	0.8750	0.1168	0.8832	0.1250	0.7750	0.1429
	(0.0165)	(0.0190)	(0.0151)	(0.0167)	(0.0120)	(0.0251)	(0.0297)	(0.0232)	(0.0427)	(0.0373)	(0.0257)	(0.0347)	(0.0236)	(0.0260)

Experimental Results: Decision Tree Structure



R

Experimental Results: Decision Tree Structure



Summary



- Collected a new RGB-D dataset (VISME) for facial micro-expressions.
- Comprehensive evaluation was performed on the VISME dataset using both classical and deep learning approaches for RGB and Depth videos on part and full face data.
- > The overall MA for the RGB videos is 1.52% higher than the depth videos.
- Depth Videos alone can be used for the classification of facial micro-expressions.



Thank You