Detecting Emergent Leader in a Meeting Environment using Nonverbal Visual Features Only

Cigdem Beyan[1], Nicolò Carissimi[1], Francesca Capozzi[2], Sebastiano Vascon[3], Matteo Bustreo[1], Antonio Pierro[4], Cristina Becchio[5,6], Vittorio Murino[1]

[1] *Pattern Analysis and Computer Vision, Istituto Italiano di Tecnologia, Genova, Italy*
[2] *Department of Psychology, McGill University, Montreal, QC, Canada*
[3] *European Center for Living Technology, Ca' Foscari University of Venice, Italy*
[4] *Department of Social and Developmental Psychology, University of Rome La Sapienza, Italy*
[5] *Department of Psychology, University of Turin, Italy*
[6] *Robotics, Brain and Cognitive Sciences, Istituto Italiano di Tecnologia, Genova, Italy*

Contact:
Cigdem.Beyan@iit.it
www.iit.it/pavis/datasets/leadershipCorpus

MOTIVATION

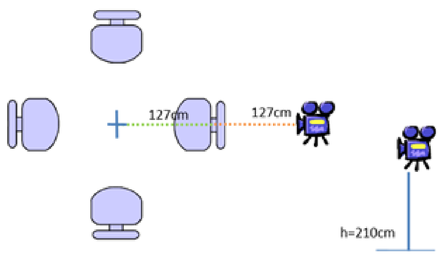
- The nonverbal visual features are not more effective compared to nonverbal audio features.
- In absence of audio sensors, the accurate detection of social interactions is still crucial.
- Better feature extraction methods and nonverbal features are needed.
- Identification of emergent leaders in small groups is a relatively new area in social signal processing.

CONTRIBUTIONS

- Devising **novel nonverbal features** based on **Visual Focus of Attention (VFOA)**.
- Introducing **a new dataset** to detect Emergent Leaders (EL).
- Presenting a **comprehensive comparison** among several **VFOA methods**.

DATASET

- 16 meetings (max. 30 minutes, min. 12 minutes).
- **Video:** 4 frontal cameras and a standard camera.
- **Audio:** 4 wireless lapel microphones.
- **Survival Task**
- **Social Psychology Questionnaires:** Multiple Level Observation of Groups (SYMLOG) & General Leader Impression Scale (GLIS).



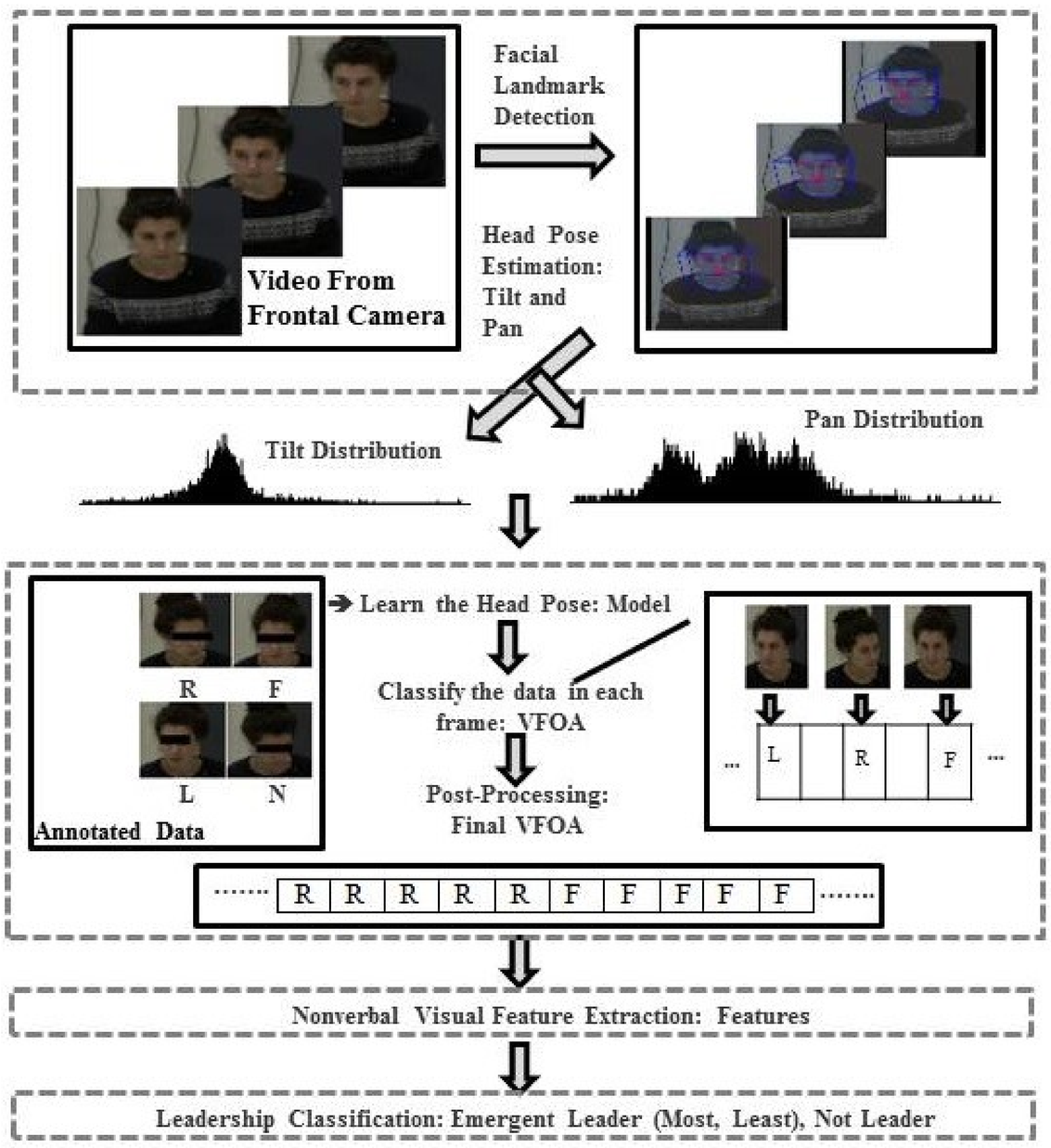
Emergent Leader	Agreement Type	Average Agreement	Total # of Meetings/ Out of
Most	Full	1	26/75
	Majority	0.73	49/75
Least	Full	1	13/75
	Majority	0.70	62/75

Data Annotation

- Meetings were divided into small segments.
- 75 meetings segments, each lasting ~5′.
- 50 annotators, each annotated 12 or 13 video segments.
- **Annotation:** The most and the least leader.

PROPOSED METHOD

- **Facial Landmark Detection and Head Pose Estimation [1]**: To extract the head pose (pan & tilt).
- **Modeling VFOA** (right, left, front, no-one)**:** SVM and its variants using 23000 labelled frames.
- **Nonverbal Visual Features:** In total 15 features. E.g.:
  - The total time that a person is mutually looking at any other persons in the meeting (total ME).
  - The total time inter-current between the initiation of ME with any other persons in the meeting.
  - The ratio between the total time that a person is being watched (total Watcher) and the total time that a person looked at other persons.
  - The maximum time that a person is looked at by any other two persons w/o ME.



RESULTS

VFOA:	R	L	F	N
OTSU	0.44	0.53	0.55	0.60
k-means	0.75	0.87	0.79	0.10
GMM	0.73	0.77	0.62	0.10
SVM	0.88	0.86	0.67	0.39
SVM-cost	0.85	0.85	0.72	0.52
SVM-RUS	0.83	0.82	0.70	0.56
SVM-SMOTE	0.87	0.86	0.70	0.51

- **The best features:** E.g. total Watcher and total ME.
- Using **the proposed features together** performed the best.
- For the most EL detection, the proposed features performed better than features in [2]. For the least EL, there was no difference.

EL Detection Rate	Most	Least	Rest
SVM	0.71	0.59	0.75
SVM-cost	0.80	0.58	0.70
SVM-PCA	0.72	0.63	0.71
SVM-PCA-cost	0.79	0.63	0.64
SVM-CorrFea	0.67	0.62	0.72
RFLA [2]	0.71	0.71	0.69
RFLA-CorrFea	0.72	0.67	0.68

Correlation Analysis:

- Except 4 features, all other features were found correlated with the SYMLOG and GLIS.

CONCLUSIONS

- The proposed nonverbal features performed well for detection of ELs (avg. detection rate ~70%).
- The majority of them were highly correlated with social psychology questionnaires.
- It is shown that accurate detection of VFOA and effective features result in more accurate EL detection.

References:
[1] T. Baltrusaitis et al., Constrained local neural fields for robust facial landmark detection in the wild, ICCVW, 2013.
[2] D. Sanchez-Cortes, Computational methods for audio-visual analysis of emergent leadership. PhD Thesis, EPFL, Lausanne, 2013.

18th ACM International Conference on Multimodal Interaction (ICMI 2016) , Tokyo, JAPAN