Moving as a Leader: Detecting Emergent Leadership in Small Groups using Body Pose

Cigdem Beyan[1], Vasiliki-Maria Katsageorgiou[1], Vittorio Murino[1,2]

[1] Pattern Analysis and Computer Vision, Istituto Italiano di Tecnologia, Via Morego 30, 16163, Genoa, Italy
[2] Department of Computer Science, University of Verona, Ca' Vignal 2, Strada Le Grazie 15, 37134, Verona, Italy

{Cigdem.Beyan,Vasiliki.Katsageorgiou,Vittorio.Murino}@iit.it

ISTITUTO ITALIANO DI TECNOLOGIA

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

MOTIVATION

- Identification of emergent leaders (ELs) in small groups is
  - A substantial topic for social and organizational psychology,
  - Becoming popular in social signal processing.
- Visual activity is a useful cue to investigate the social interactions.
  - Previous works [1,2]: head /body activity (*Head/Body_Act*)
  - Not performing well enough.

CONTRIBUTIONS

- Introducing novel visual activity-based nonverbal features (NF) using 2D-body pose.
  - First time: Body pose used for leadership (ELship) detection.
- Investigating the effect of unsupervised feature learning.
  - First time: this is applied for ELship.
- Improved performance as compared to SoA of visual activity-based NFs [1,2].

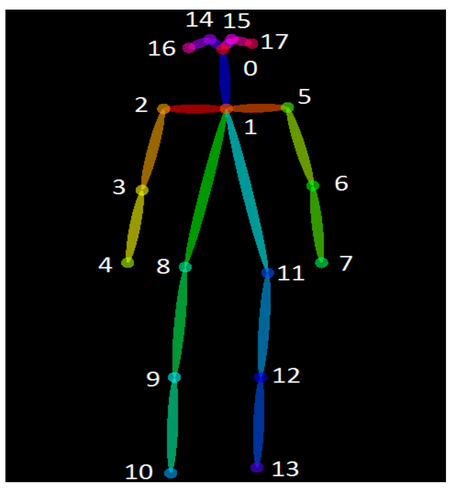
DATASET

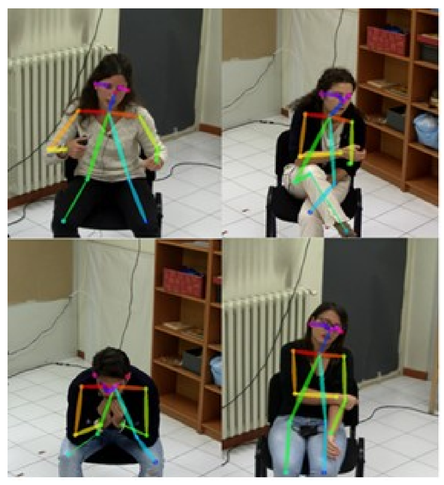
- 16 meetings (12-30 minutes).
- **Video:** 4 frontal cameras and a standard camera.
- **Audio:** 4 wireless lapel microphones.
- **Survival Task**
- **Annotation:** The most & least emergent leaders.
- 75 meeting segments, 300 samples in total.

Publicly Available

PROPOSED METHOD

1) 2D-Body Pose Estimation [3]
- Based on Convolutional Neural Networks.
- Applied to the frames having significant activity.
- 12 body parts used (i.e. ears, eyes, feet discarded).
- Missing body parts: Spline interpolation.





2) Nonverbal Feature Extraction
- 10 angles between body parts: the cosine value of the two vectors.
- 8 statistical measures: mean, standard deviation, skewness, the number of zero crossings, etc.

3) Classification: the most EL, the least EL, not-a-leader
- Support Vector Machines (SVM)
- Localized Multiple Kernel Learning (LMKL) [4]

Detection Rate	Most EL	Least EL	Not-a-leader
Head/Body_Act_SVM	**0.57**	0.36	**0.70**
Pose_SVM (Proposed)	0.46	**0.54**	0.55
Head/Body_Act_LMKL	0.52	0.42	**0.94**
Pose_LMKL (Proposed)	**0.54**	**0.46**	0.92

- Overall, the *proposed body pose-based NFs* performed better than existing visual activity-based NFs.
- It is possible to improve classification results by applying unsupervised feature learning (DBMs).
- The *proposed NFs* were able to advance the EL identification performances of other types of NFs when they were used together.
  - *VFOA_Pose* performed better than *VFOA* only and *VFOA_Head/Body_Act*.
  - Any fusion w/*SpeakAct* performed worse than *SpeakAct* only.

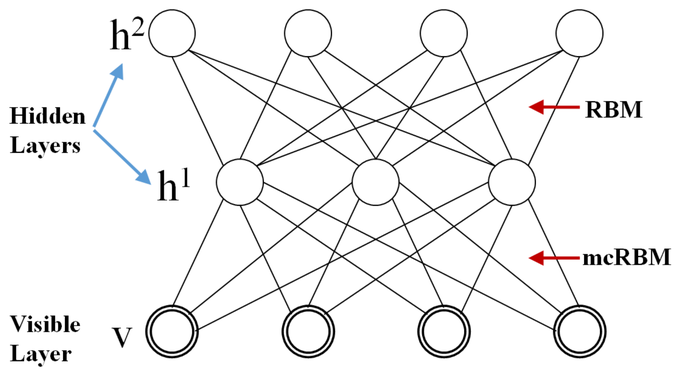
  BUT
  - *VFOA_Pose* (also best of all) performed better than *SpeakAct* and any combinations with *SpeakAct*.

RESEARCH QUESTIONS

1) Are the *Proposed pose-based NFs* **better than** *Head/Body_Act*?

Feature Learning:

2) Are the *Proposed NFs* w/Deep Boltzmann Machines (DBMs) **better than** *Proposed* only?
   - Are the *Proposed NFs* w/DBMs **better than** *Head/Body_Act* w/DBMs?



Fusion with other NFs:

3) Are the fusions of *proposed NFs* w/speaking activity-based NFs (*SpeakAct*) and visual focus of attention-based NFs (*VFOA*) **better than** *SpeakAct* only and *VFOA* only?
   - Are the fusions of *proposed NFs* w/other NFs **better than** fusion of *Head/Body_Act* w/other NFs?

Detection Rate	Most EL	Least EL	Not-a-leader
Head/Body_Act_DBM_SVM	0.52	0.30	**0.65**
Pose_DBM_SVM (Proposed)	**0.54**	**0.48**	0.52
Head/Body_Act_DBM_LMKL	**0.68**	0.27	0.85
Pose_DBM_LMKL (Proposed)	0.55	**0.50**	**0.94**

Detection Rate	Most EL	Least EL	Not-a-leader
SpeakAct_SVM	**0.73**	**0.59**	0.73
SpeakAct_Head/Body_Act_SVM	0.69	0.52	0.74
SpeackAct_Pose_SVM (Proposed)	0.71	0.52	**0.76**

Detection Rate	Most EL	Least EL	Not-a-leader
VFOA_SVM	0.71	0.59	0.75
VFOA_Head/Body_Act_SVM	0.70	0.64	**0.78**
VFOA_Pose_SVM (Proposed)	**0.72**	**0.66**	0.74

References:
[1] C. Beyan et al., Identification of Emergent Leaders in a Meeting Scenario Using Multiple Kernel Learning, ACM ICMI-ASSP4MI, 2016.
[2] D. Sanchez-Cortes et al., A Nonverbal Behavior Approach to Identify Emergent Leaders in Small Groups, IEEE Trans. Multimedia, 2012.
[3] Z. Cao et al., Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. CoRR abs/1611.08050, 2016.
[4] M. Gonen et al., Localized Multiple Kernel Learning, ICML, 2008.