Towards a Method to Detect F-formations in Real-Time to Enable Social Robots to Join Groups

Sai Krishna
Center for Applied Autonomous Sensor Systems (AASS), Orebro University
Orebro, Sweden
sai.krishna@oru.se

Andrey Kiselev
Center for Applied Autonomous Sensor Systems (AASS), Orebro University
Orebro, Sweden
andrey.kiselev@oru.se

Amy Loutfi
Center for Applied Autonomous Sensor Systems (AASS), Orebro University
Orebro, Sweden
amy.loutfi@oru.se

ABSTRACT

In this paper, we extend an algorithm to detect constraint based F-formations for a telepresence robot and also consider the situation when the robot is in motion. The proposed algorithm is computationally inexpensive, uses an egocentric (first-person) vision, low memory, low quality vision settings and also works in real time which is explicitly designed for a mobile robot. The proposed approach is a first step advancing in the direction of automatically detecting F-formations for the robotics community.

CCS CONCEPTS

• Human-centered computing → Human computer interaction (HCI); **• Computing methodologies → Computer vision**; **Vision for robotics**;

KEYWORDS

Social Robot, F-formations, Face Orientation

ACM Reference Format:

Sai Krishna, Andrey Kiselev, and Amy Loutfi. 2017. Towards a Method to Detect F-formations in Real-Time to Enable Social Robots to Join Groups. In *Proceedings of ECCE Workshop, Umea, Sweden, September 2017 (ECCE '17),* 2 pages.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Humans organise themselves spatially while interacting with others (social interaction). As social robots are increasingly entering our house and work space, it is of equal importance that these systems promote conditions for good social interaction by exploiting the natural spatial interactions that humans use. The spatial orientations that humans use have been described by Adam Kendon's [1] Facing formations, known as F-formations, which are spatial and orientational relationship between two or more people conversing with each other. Kendon proposed four standard formations. *Vis-a-Vis* is when two people are interacting while facing each other. *Side-by-Side* is when two people stand close to each other and face the same direction while conversing. *L-shape* is when two people are facing perpendicular to each other and are placed on the two edges of letter 'L'. When three or more people are conversing in a circular arrangement then it is *a circular formation.*

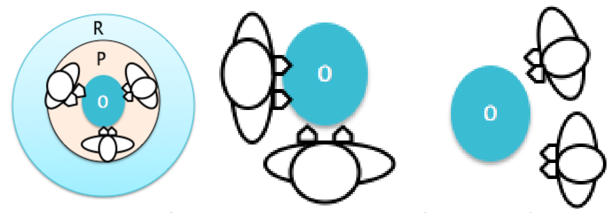


Figure 1: ***An F-formation gives rise to three social spaces. O-Space: Convex empty space, P-Space: narrow strip on which people are standing & R-Space: beyond p-space. Left:Circular, center:L-Shape, right:Side-by-Side***

Many researchers proposed different methods to detect F-formations. Few of them are: Cristani et al [2] used Hough voting strategy to locate the o-space, Graph-Cuts for F-formation (GCFF) [3] detects groups in still images using proxemic information, Vascon et al [4] generate a frustum based on the position and orientation of person and compute affinity to extract F-formation. In [5], the authors consider the body orientation as the primary cue and propose a joint learning approach to estimate the pose and F-formation for groups in videos, these methods are from computer vision community.

Vazquez et al [6] explored this problem in robotics community, proposed detecting the F-formations based on lower body estimation but uses an exocentric camera (overhead video data set). So, yet there is an unsolved problem of detecting F-formations in real time on a mobile robot. Our paper [7] explored this topic and detected the F-formations based on the face orientation. The algorithm proposed works in real time, uses low memory, low quality vision settings, computationally inexpensive and uses an egocentric (first-person) vision but yet there are few constraints: works in laboratory settings, constraint based formations and the situation when the robot is in motion were not addressed. Constraint based formations: when one person is facing two or more people while interacting it is a *Triangle formation*, for example, the

ECCE '17, September 2017, Umea, Sweden

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

ticket counters. *Rectangle formation* is formed in board meeting rooms and *Semi-Circular formation* is formed when three or more people are focusing on same task while interacting with each other, for example, in-front of a wall while watching a piece of art.

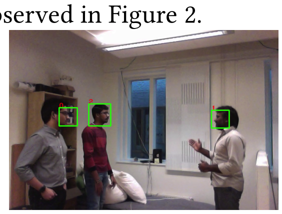
In this paper, we extend our algorithm with few more hypothesis to detect constraint based F-formation and also consider the situation when the robot is in motion.

2 METHODOLOGY

Our method [7] uses Haar cascade face detector algorithm to detect the faces of the people in the scene. Once faces are detected, we identify the location of the eyes, to obtain rough estimation of the orientation of the person.

The methodology is based on quadrants. In our model, we assume, if a person is looking towards right then both eyes can be located in first quadrant; if the person is looking towards left then both eyes can be located in second quadrant; if looking towards center then both eyes are in both the quadrants, one in each. This is obviously a very rough approximation but we are interested only in the facing direction of the person and not in calculating the exact angle of the head. After estimating the facing direction, the F-formations are detected by mapping these face orientations to the spatial arrangements depending on the number of people.

We extend this algorithm to account, when only one eye is visible. If a person's left eye is located in first quadrant and no eyes in the second quadrant then we assume that they are facing left and if a person's right eye is located in second quadrant and no eyes in the first quadrant then they are facing right which can be seen in Figures 3 & 2. Using these hypothesis, we detect triangle formation which is a constraint based F- formation and arises when one person is facing left and two persons are facing right or vice versa which can be observed in Figure 2.



(a) Triangle

Figure 2: *F-formations in real time.*

3 EXPERIMENTAL RESULTS

We performed qualitative experiments to estimate the F-formations in real time with a webcam of 640 x 480 pixel resolution on a standard PC and tested on 11 people with different spatial configurations. The results obtained are at a rate of 20 fps. The experiment scenario was done in this way: a camera (robot's vision) is placed at a height of 1.5m, people walk into the scene (in front of camera) and have a natural conversation. The algorithm starts detecting people's faces, estimates their face orientation and then identifies the F-formations. The algorithm works as far as 2.5 meters between the people & the robot. We also tested the algorithm by moving the camera left to right and right to left. Assuming the robot has 1 DOF, which is the *yaw axis* similar to humans moving their head side to side. The results can be observed in Figure 3.

Figure 3: *Frames:when robot (camera) is in motion.*

4 CONCLUSIONS

In this paper, we extended the proposed method by few more assumptions and also, tested the algorithm when the robot is in motion. In future, we would extend this algorithm to other formations & natural settings and also create a dataset to evaluate the algorithms in real time.

REFERENCES

[1] A. Kendon, "Spacing and orientation in co-present interaction," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5967 LNCS, 2010, pp. 1–15.

[2] M. Cristani, L. Bazzani, G. Paggetti, A. Fossati, D. Tosato, A. D. Bue, G. Menegaz, V. Murino, Andrea Fossati, and A. Del˜Bue, "Social interaction discovery by statistical analysis of F-formations," *British Machine Vision Conference (BMVC)*, 2011.

[3] F. Setti, C. Russell, C. Bassetti, and M. Cristani, "F-formation detection: Individuating free-standing conversational groups in images," *PLoS ONE*, vol. 10, no. 5, 2015.

[4] S. Vascon, E. Z. Mequanint, M. Cristani, H. Hung, M. Pelillo, and V. Murino, "A Game-Theoretic Probabilistic Approach For Detecting Conversational Groups," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9007, 2015, pp. 658–675.

[5] E. Ricci, J. Varadarajan, R. Subramanian, S. R. Bulo, N. Ahuja, and O. Lanz, "Uncovering interactions and interactors: Joint estimation of head, body orientation and f-formations from surveillance videos," in *Proceedings of the IEEE International Conference on Computer Vision*, vol. 11-18-Dece, 2016, pp. 4660–4668.

[6] M. Vázquez, A. Steinfeld, and S. E. Hudson, "Parallel detection of conversational groups of free-standing people and tracking of their lower-body orientation," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on.* IEEE, 2015, pp. 3010–3017.

[7] S. K. Pathi, A. Kiselev, and A. Loutfi, "Estimating f-formations for mobile robotic telepresence," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction.* ACM, 2017, pp. 255–256.