Social Groups Detection in Crowd by Using Automatic Fuzzy Clustering with PSO

Ali Akbari
Department of Electrical and Computer Engineering., University of Birjand, Birjand, Iran.
ali.akbari@birjand.ac.ir
Hassan Farsi*
Department of Electrical and Computer Engineering., University of Birjand, Birjand, Iran.
hfarsi@birjand.ac.ir
Sajad Mohamadzadeh
Technical faculty of Ferdows, University of Birjand, Birjand, Iran.
s.mohamadzadeh @birjand.ac.ir

Received: 04/Sep/2019 Revised: 24/Nov/2019 Accepted: 08/Dec/2019

Abstract

Detecting social groups is one of the most important and complex problems which has been concerned recently. This process and relation between members in the groups are necessary for human-like robots shortly. Moving in a group means to be a subsystem in the group. In other words, a group containing two or more persons can be considered to be in the same direction of movement with the same speed of movement. All datasets contain some information about trajectories and labels of the members. The aim is to detect social groups containing two or more persons or detecting the individual motion of a person. For detecting social groups in the proposed method, automatic fuzzy clustering with Particle Swarm Optimization (PSO) is used. The automatic fuzzy clustering with the PSO introduced in the proposed method does not need to know the number of groups. At first, the locations of all people in frequent frames are detected and the average of locations is given to automatic fuzzy clustering with the PSO. The proposed method provides reliable results in valid datasets. The proposed method is compared with a method that provides better results while needs training data for the training step, but the proposed method does not require training at all. This characteristic of the proposed method increases the ability of its implementation for robots. The indexing results show that the proposed method can automatically find social groups without accessing the number of groups and requiring training data at all.

Keywords: Author Guide; Article; Camera-Ready Format; Paper Specifications; Paper Submission.

1- Introduction

Over the time and the extending use of a camera to maintain security and detect social anomalies, the importance of the processing video data has increased. Detecting social groups is concerned by governments for detecting dangerous situations and analyzing people's behavior [1]. Detection of theft and terrorist groups is of great importance. To identify groups with such aims, recognizing social groups is a prerequisite. In [2], social anomalies between two persons were analyzed.

According to the recent research in [3], people are interested in moving in social groups of crowds and the behavior of pedestrians in social groups of the crowd has been studied. According to the approach in [4], people are interested in moving in groups. Moving in the group ~~means to be a subsystem in th~~e group, in other words, a group containing two or more persons can be considered in terms of the same direction of movement and speed of movement. According to the studies in [5], most social groups contain two persons and social groups including three and four persons are at top of the list in terms of people's willingness to appear in these groups and to move in crowds. According to the research in [6], a group is considered by the personal influence of someone or other people to move through crowds. The person who enters into a group is influenced by the group, and the speed and direction of that person are changed and grouped accordingly. Although many methods for the detection of social groups have been reported so far, all of them need training data. A method that does not need training data has not been reported yet.

This paper is organized as follows: In Section 2, a literature review is explained. In Section 3, the physical distance feature and automatic fuzzy clustering with the PSO are introduced and the proposed method is detailed for identifying the social groups. In Section 4,

* Corresponding Author:

experimental results are demonstrated and in the final section, the conclusion is drawn.

2- Literature Review

For detecting social groups several methods have been reported so far. Some of the methods used many features and machine learning. These methods require training their networks. Some of the methods used people's movement and these methods are divided into three categories. Some new methods have been used deep learning, deep neural network and Long Short Term Memory (LSTM) in recent years.

In [7], the detection of social groups to help the robot's behavior in teamwork with humans has been reported, and an anticipation method by linear extrapolation of inter-event intervals has been used. In [8], social groups have been detected by skeletal data from participants, and the event anticipation method has been used to move robots among human groups. The method reported in [9] detects conversation and social groups for the robot by using the direction of people and the lower-body orientation. In [10], body, head orientation and body orientation from a distance social scene are used to detect social groups. As an example, the output of this method is shown in Figure 1.

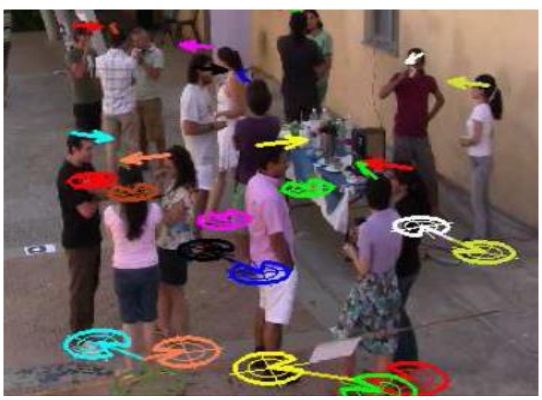


Fig. 1. head orientation and body orientation used in [10].

In [11], social groups are detected by using physical distance, people's locations and head directions in definite spaces. In this method, fixed and unmoving groups are detected. Figure 2 shows the various models concerned by this method.

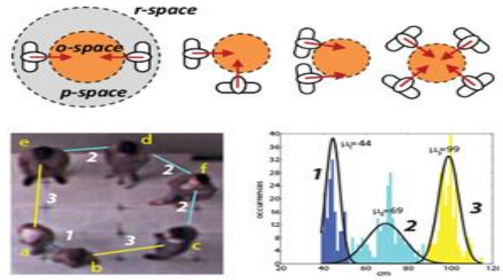


Fig. 2. Verity of direction in definite spaces [11].

Social group detection based on direction of people's movement is divided into three categories: group-based, individual-group joint and individual-based. The reported method in [12] applies the second-order derivative of the information matrix of people's movement path to detect social groups.

Individual-group joint approaches use more important information compared to group-based approaches, for instance, group tracking [13]. In individual-based approaches, the groups are detected by using single people's trajectories. As an example, in [14], the head is more accurately detected using crowd density estimation.

Fig. 3. Effect of crowd density estimation in head detection in a crowded area.

The reported method in [15] predicts two persons in one group using distance features, speed difference, and time overlap information with Support Vector Machines (SVM).

In [16], probabilistic similarity methods of two-persons variations in successive frames and soft areas are used to identify social groups.

An attractive method without training is introduced in [17], which uses group information in the previous frame. In this method, a potentially infinite mixture model of group probability with the mean value of distance in burst frames is examined. In this method, only the distance feature is used and, better results are achieved in the recognition of social groups by using more features.

In [18], proximity and velocity characteristics are used to identify social groups. In this method, in crowded conditions, Markov model, proximity and velocity features are assisted, and in some situations which are not very crowded, the pursuit of individuals in successive frames is determined to recognize the social groups.

In Figure 4, the status of people is shown in sequential frames.

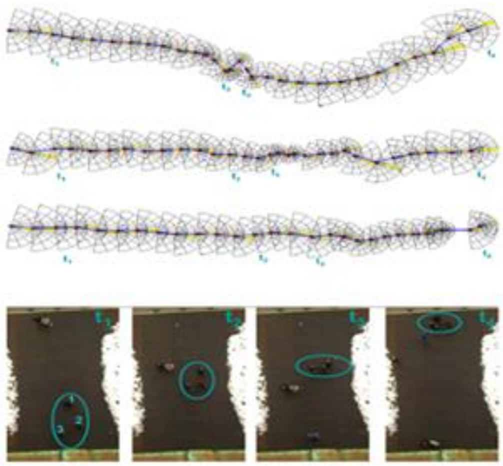


Fig. 4. Tracking people in consecutive frames to identify social groups [18].

The reported method in [18] has detected social groups by using the features of distance and attention to people or the direction of the people in sequences of frames and game theory tools.

In the method reported in [19], which is inspired by electric dipole shown in Figure 5, each person's eyesight is firstly examined, and if a relationship between the attention of the people's eyes is found, this group of people is placed in a social group.

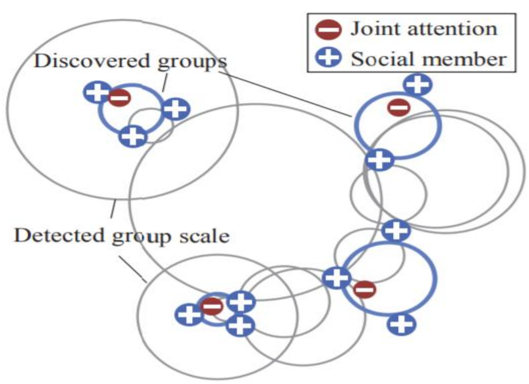


Fig. 5. Effect of investigating the relationship between visual attention in the detection of social groups [19].

In the reported method in [20], the relationship between people is identified by using graph-based clustering and further social group activity is detected by the SVM-based classification.

In [21], the track of salient points and adaptive clustering are used, and hierarchical social groups are achieved. In the method reported in [22], by weighing two features including distance and speed correlation between two persons, social groups are detected, and the group's behavior is analyzed. In [23], instead of using the similarity feature between two persons, the clustering of pedestrians into different groups is based on using the start and endpoint to identify social groups. In further analysis of square matrices, the size of the number of people in the video is constructed and the probability of being grouped is calculated between all of them. The probabilities are calculated based on Euclidean distance between two individuals. Hierarchical clustering is then used to identify social groups. In [24], conjoint individuals and group tracking have been reported. RGB histogram, region covariance, and Histogram Of Gradient (HOG) similarity are used to detect social groups.

In [25], features of distance, motion causality, trajectory shape, and path convergence are used by optimized the SVM to detect social groups. The reported method in [26] has used deep learning algorithms, trajectory modeling approaches with LSTM, contextual information from the local neighborhood and Generative Adversarial Network (GAN) to detect social groups.

3- Proposed Method

In the proposed method, automatic fuzzy clustering with the PSO has been used to cluster all people in the video. After automatic fuzzy clustering with the PSO, groups are detected. Post-processing is used to remove scattered groups. After the post-processing, final social groups are maintained. In each iteration, the automatic algorithm merges two similar clusters and the fuzzy PSO changes 'w' to approach a global solution. In the following, the proposed method is discussed in four main parts as shown in Figures 6 and 7.

```
Algorithm:
1: Input = people's  location in video
2: k = number of people in video // only for the first iteration
3:   for iteration < max iteration
4:        clustering by PSO for k clusters
5:       S // calculate scattering for all clusters (calculate using
//equation (4))
6:       R // calculate similarity between two clusters (calculate
//using equation (6))
7:           if (max (similarity < 2) and similarity>0.4) between
two // clusters
8:           make one cluster with them and k = k-1
9:           end if
10:     w // fuzzy operation identified weight for PSO
11:   end for
12: post-processing // delete this group (maximum distance
//between group members from the center of the group >0.4)
13: output = social groups detection
```

Fig. 6. *Algorithm: the proposed method.*

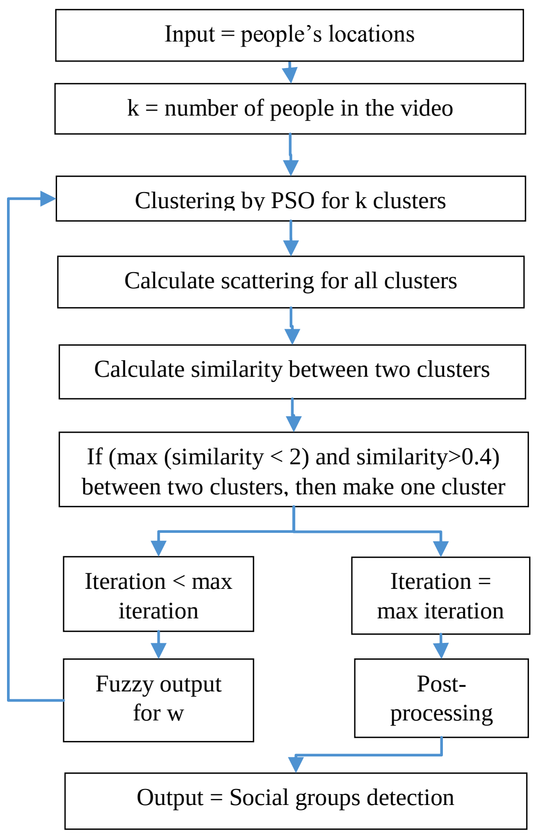


Fig. 7. Flowchart of the proposed method.

In the following, all parts of the flowchart depicted in Figure 7 are described. The physical distance feature is used for clustering by calculation of scattering and similarity.

3-1- Physical Distance Feature

One of the most important features in detecting social groups is physical distance between people in a scene. The proxemics theory is identified based on physical distance features [27]. This theory is shown in Table 1.

Table 1. The proxemics Theory [27]

space	Boundaries(m)	description
Intimate	0.0 – 0.5	Unmistakable involvement
Personal	0.5 – 1.2	Familiar interactions
Social	1.2 – 3.7	Formal relationships
public	3.7 – 7.6	Non-personal interactions

The proxemics theory includes 4 classes to identify the relation between pairs and their physical distance. Each class is separated from another class by the boundaries shown in Table 1.

3-2- Random Algorithm

In [28], a randomized algorithm, with the PSO has been reported based on collective intelligence. In this algorithm, random particles are firstly generated. Next, all the particles are applied to a fitness function and the best solution among all solutions, P_g^t , is the leader of all particles. For all particles, the best position of the particles is saved as the leader of the P_i^t particle. In Equation 1, Y_i^t is the position of the i-th particle at the moment 't', and V_i^{t+1} is the velocity of the particle [28].

$$V_i^{t+1} = w * V_i^t + c_1 * r_1 * \left(P_g^t - Y_i^t\right) + \\ + c_2 * r_2 * (P_i^t - Y_i^t) \tag{1}$$

In this equation, $r_1$and r_2 are random numbers between zero and one, and c_1 and c_2 are considered to be 2 for the PSO algorithm [28]. The speed of each particle is updated at each iteration. In Equation 1, 'w' is the setting parameter for exploiting or exploring the search space. If 'w' is large, the speed will be higher and the search step will increase. This search for the problem space is called exploration. If w is small, the speed will reduce and the search step will be short. This type of searching is called exploitation. Next, according to Equation 2, the positions of all particles are updated [28].

$$Y_i^{t+1} = Y_i^t + V_i^{t+1} \tag{2}$$

In Equation 3, 't' is the current repetition and 'maxt' is the maximum iteration to get the best solution in the search space [28].

$$w = \frac{-0.8 * t}{maxt} + 0.9 \tag{3}$$

3-3- Clustering

For a more precise clustering, the Davis-Bouldin (DB) index is used [29].

3-3-1- Davis-Bouldin Index

Different measures are used for clustering. One of the most important measures is the Davis-Bouldin index [30, 31]. For a more precise clustering, the distance between cluster members and the center of their clusters should be minimized, and the cluster members' distance between the other clusters should be maximized. In the following, the formulation of this concept is discussed. Scattering of a cluster, S_i, is given by [30]:

$$S_i = \left(\frac{1}{T}\sum_{j=1}^{T_i} |X_j - A_i|^p\right)^{\frac{1}{p}} \tag{4}$$

where A_i is the cluster center and 'T' is the number of members located in the i-th cluster. If p is equal to 2, then Euclidean distance is calculated between the cluster members and the cluster centers.

The disparity between the two clusters is defined as follows [30]:

$$d_{ij} = \|A_i - A_j\|_p = \left(\sum_{k=1}^{n} |a_{k,i} - a_{k,j}|^p\right)^{\frac{1}{p}} \quad (5)$$

Where k is the dimension for the center in the cluster. For both equations, p is considered to be 2.

The concept of similarity between two clusters is defined as [30]:

$$R_{ij} = \frac{S_i + S_j}{d_{ij}} \quad (6)$$

The similarity between the two clusters, R_{ij}, has the following properties:
$R_{ij} \geq 0$
$R_{ij} = R_{ji}$
If S_i and S_j both are zero, then R_{ij} will be zero.
If $S_j \geq S_b$ and $d_{ij} = d_{ib}$ then $R_{ij} > R_{ib}$
If $S_j = S_b$ and $d_{ij} \leq d_{ib}$ then $R_{ij} > R_{ib}$

The DB index is now defined as [29]:

$$DB = \frac{1}{N}\sum_{i=1}^{N} D_i \quad , D_i = \max_{j \neq i} R_{ij} \quad (7)$$

where N is the number of clusters. By considering the maximization of the similarity between two clusters, since the distance between the two centers is in the denominator of the similarity between two clusters, R_{ij}, it is needed that the distance between the two centers of similar clusters to be minimized. This means that two similar groups are considered in the same group. Of course, by maximizing the similarity between two clusters, the, DB index will also be larger. As a result, with a larger DB index, the clustering can detect better clusters.

3-3-2 Automatic Clustering

Automatic clustering means that the number of clusters is unknown, as an unsupervised, and the algorithm must recognize the number of clusters and perform clustering [33, 34]. For this aim, in the process of finding the similarity between two clusters, the groups with the similarity between two clusters under the threshold of 0.4 are removed. Also, if the sum of the similarities between the two clusters is less than 2, then two groups with the highest similarity between two clusters are merged. This means that the number of groups is reduced and, of course, the condition of being less than 2 is to reduce the number of clusters at first when the number of groups is large. Then reducing the number of clusters will be stopped when approaching the correct numbers of the clusters.

3-3-3- Automatic Clustering with PSO

For all clusters, PSO randomly specifies cluster centers. In the following, with automatic clustering, the number of clusters is reduced, but ‘w’ and the particle size are not reduced. This means that in each iteration all centers are updated by the PSO, but the cluster centers that satisfy the automatic clustering conditions will be entered into the next processing stage. In the initial implementation, in the proposed method, the number of clusters per video is equal to the number of people existing in the same video. In the following, the cluster centers of each video are randomly identified by the PSO.

3-3-4- Automatic Fuzzy Clustering with PSO

PSO is a relatively recent heuristic search method which is based on the idea of collaborative behavior and swarming in biological populations. The PSO senses population-based search approaches and depending on information sharing among their population members enhances their search processes using a combination of deterministic and probabilistic rules. The weakness of the PSO is detecting local solutions and not approaching a global solution when it solves with fuzzy PSO.

In the PSO algorithm, the weight ‘w’ decreases when iteration increases and reaches to its minimum in final iterations. Decreasing ‘w’ means that the search is around the previous solution ‘s’. This also means that the PSO finds the approximate location of the solution s and decreases ‘w’ to find the exact location around the current location. However, the problem is that it may not be close to the proper solution. To solve this problem, fuzzy controllers are used. It is even possible to use fuzzy controllers for an individual and group leader's impact factor in the PSO algorithm. In the proposed method, fuzzy control is used only for ‘w’. In each execution, the best scattering solution of a cluster S_i is examined. If the best dispersion solution among all clusters is very low, ‘w’ is considered small, which means we are close to the global solution, and we need to perform a thorough review around it. Now, if the best solution among all clusters is a large number, then we have a far-reaching global solution, and we need to search the global solution with a large ‘w’ within the entire search space.

3-4- Post-processing

The output includes clustering with clusters far from the center of the cluster and also a short distance from the center of the cluster, but long-distance clusters are not suitable for the detection of social groups. As a result, a threshold for maximum distance between group members from the center of the group is considered and the threshold in the proposed method is 0.4. The first video among twenty videos in student03 database contains 1475 frames. These frames are converted into 25 episodes and the average position of the people is recorded. Due to the difference in the number of frames and therefore the average positions of the people (for example in the twentieth video, there are five average positions of the people) in the proposed method we have converted all videos into five average positions of individuals in equal intervals. In each of the five average positions of the individuals per video, 400 replications are performed by the PSO for automatic fuzzy clustering.

Among all parts of each video the positions of all people are averaged as shown in Figure 8. This means that the positions of all people in groups that move from one part of an image to another part are placed near them.

Fig. 8. Few frames from the first set of the first video

As shown in Figure 9, the convergence is achieved by the PSO after 400 iterations for one part of the five episodes of a video. After convergence, the post-processing is started. As shown in Figure 10, the post-processing neglects the groups with high dispersion. In Figure 10, circles are drowned in the center of the groups.

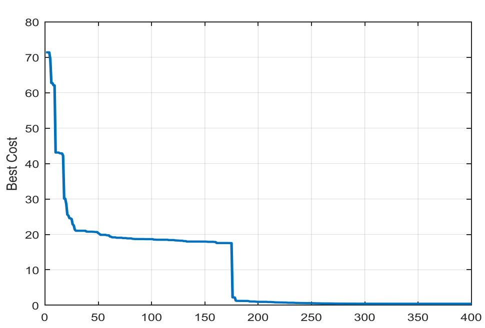


Fig. 9. Automatic fuzzy clustering convergence by the PSO for 400 iterations for one part of the five episodes of a video.

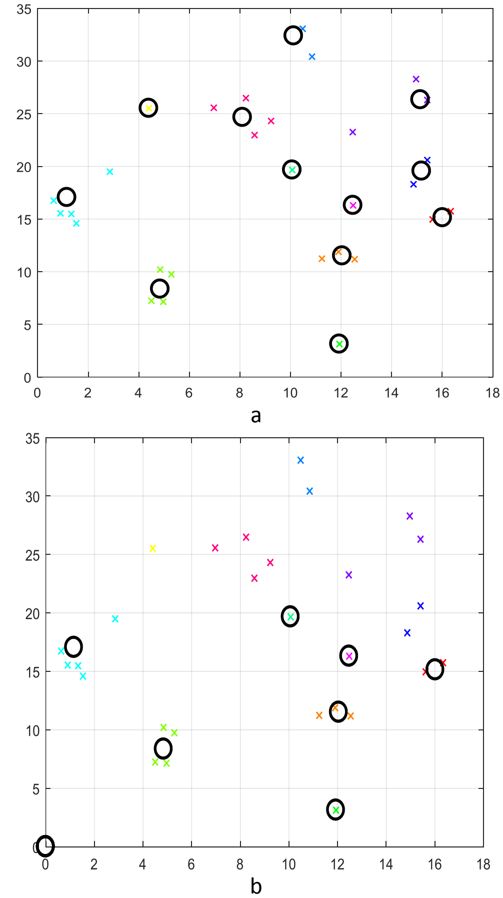


Fig. 10. (a) Automatic fuzzy clustering output by the PSO, (b) Automatic fuzzy clustering by PSO after post-processing and removal of high-dispersion clusters.

4- Experimental Results

4-1- Datasets

To evaluate the performance of the proposed method, ETH and Hotel [33] ,Student003 [25], GVEII [26, 35] and MPT-20x100 [25] datasets have been used. These datasets include the trajectory of people and the number of people. As an example in Figure 11, the trajectory and the number of people are shown in a sequence of frames.

Fig. 11. The trajectory of people's movement and the number of people in the sequence of frames.

In Table 2, five datasets are compared. The parameters v, p, g, d1, and d2 represent the number of videos, the number of people, the number of groups, the minimum distance in the group (m), and the maximum distance in the group (m), respectively.

Table 2. Comparing four datasets.

	v	p	g	d1	d2
ETH	1	117	18	0.99	2.79
Hotel	1	107	11	0.75	2.00
Student003	20	406	108	0.41	0.71
GVEII	30	630	207	0.77	1.66
MPT-20x100	20	82	10	0.63	1.45

4-2- Evaluation Measures

The precision and recall parameters are used to compare the results. The precision is the ratio of the number of groups that are correctly identified to the number of all identified groups; the recall is the ratio of the number of groups that are correctly identified to the number of real groups of the database. The standard F-score, F1, is defined as follow [33]:

$$F1 = \frac{2 * \text{precision} * \text{recall}}{\text{precision } + \text{recall}} \quad (8)$$

4-3- Indexing Results

Since ETH, Hotel, student003, GVEII and MPT-20x100 datasets have different properties, they are used to evaluate real scenarios. In this section, we present the obtained results of the proposed method and compare it with other methods. The precision and recall parameters are used to compare the proposed method with social constrained structural learning (SCSL) for group detection in the crowd [25], vision-based analysis of small groups (VASG) in pedestrian crowds [18], who are you with and where are you going? (WWG) [15], online Bayesian non-parametric (OBNP) for group detection [17], scene-independent group profiling in the crowd (SIGP) [12], conjoint individual and group tracking (CIGT) [24] and Generative Adversarial Networks for Trajectory Prediction and Group Detection in Crowds (GD-GAN) [26].

Table 3. The obtained results in ETH and Hotel datasets.

	ETH		Hotel	
	precision	recall	precision	recall
proposed method	90.56 ±0.34	**86.79** ±0.42	**89.23** ±0.67	**91.81** ±0.26
CIGT [24]	**96.24**	56.16	-	-
SCSL [25]	91.14	83.43	89.12	91.32
VASG [18]	80.72	80.78	88.93	89.35
WWG [15]	60.64	76.42	84.06	51.27
OBNP [17]	79.12	80.76	88.12	73.25
SIGP [12]	69.33	68.26	67.38	64.11

The precision and recall are calculated and presented in Table 3 for the ETH and the Hotel datasets. Each measure is reported in terms of mean and standard deviation over 5 runs to account for the stochastic nature of the clustering of our algorithm. Although the proposed method is not needed to be trained, the recall of the proposed method in the ETH dataset is 86.79 and the recall of the proposed method in the Hotel dataset is 91.81, which is the best between all methods. The recall of the SCSL method in the ETH dataset is 83.43 and the recall of SCSL in the Hotel dataset is 91.32. The precision of the CIGT and the SCSL methods in the ETH dataset are 96.24 and 91.14, respectively, where the precision of the proposed method is 90.56.

The precision of the proposed method in the Hotel dataset is 89.23, which is the best performance among other methods. The precision of the SCSL method in the Hotel dataset is 89.12.

To find the precision and recall for all 20 videos, 20 videos of the Student003 dataset and 30 videos of the GVEII dataset are calculated separately in five sections, and the averages of these five measures are compared. Tables 4 and 5 present the average of all sections for the student003 and the GVEII datasets, respectively. Each measure is reported in terms of mean and standard deviation over 5 runs to account for the stochastic nature of the clustering of the proposed algorithm for the student003 and the GVEII datasets.

Table 4. Obtained precision and recall for the student003 dataset

	Precision	recall
proposed method	80.78 ±0.36	**91.81** ±0.21
SCSL	81.72	82.51
GD-GAN	82.14	63.47
VASG	77.28	73.69
WWG	56.76	76.02
OBNP	71.12	78.76
SIGP	40.48	48.63

The results of the same experiment for student003 dataset are summarized in Table 4. The recall of the proposed method in the student003 dataset is 91.81, which has the best performance. The results of the same experiment for the GVEII dataset are summarized in Table 5. The recall of the proposed method in GVEII dataset is 96.31, which has the best performance.

Table 5. Obtained precision and recall for the GVEII database

	precision	recall
proposed method	**85.53** **±0.24**	**96.31** **±0.14**
SCSL	84.12	84.11
GD-GAN	83.16	79.54
VASG	80.14	79.45
WWG	57.84	75.51
OBNP	70.15	76.65

SIGP	44.85	49.92

As shown in Tables 4 and 5, the recall of the proposed method in the student003 and the GVEII datasets is the best among all methods. Note that the proposed method is not needed to be trained.

Some of the videos in the MPT-20x100 database are selected and F-score is calculated for the proposed method. Table 6 presents the F-score of the proposed method in each video, and compare it with the SCSL and the VASL methods. The results of the same experiment for the MPT-20x100 dataset are summarized in Table 6. As shown in Table 6, F-scores of the proposed method are 79.38, 97.86, 98.47, and 96.19 for the airport1, chinacross4, grand1, and thu10 videos, respectively; these are the highest scores compared to other methods.

Table 6 Results of comparison in MPT-20x100 database

F-score	airport1	china cross4	grand1	thu10
proposed method	**79.38**	**97.86**	**98.47**	**96.19**
SCSL	78	96	97	90
VASG	58	92	83	82

Effect of the threshold in the post-processing is examined, and the precision and recall are determined with the threshold of 0.3, 0.4 and 0.5 for the student003 and the GEVII datasets. The obtained results are presented in Tables 8 and 9. Note that in all previous Tables the threshold is 0.4.

Table 8. Precision and recall for different thresholds in student003 dataset.

Threshold=0.3		Threshold=0.4		Threshold=0.5	
precision	recall	precision	recall	precision	recall
64.25	92.35	64.38	91.81	62.43	90.13

Table 9. Comparison of thresholds in the GEVII dataset

Threshold=0.3		Threshold=0.4		Threshold=0.5	
precision	recall	precision	recall	precision	recall
85.32	97.11	84.48	96.31	80.97	92.69

As observed, the best performance is provided by a threshold of 0.3. This means that removing larger groups results in better precision and recall.

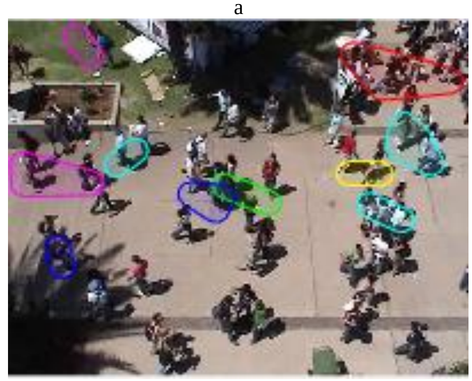
As an example, Figures 12 and 13 show detecting social groups by the proposed method.

5- Conclusion

Detecting social groups is one of the most important problems that has been concerned recently to analysis interpersonal relations in groups. In this study, an automatic fuzzy clustering with the PSO was used, and acceptable results were achieved.

It is a matter of great importance, without supervision and training, of the automatic fuzzy clustering with the PSO. This method does not require to be trained and is easier to be calculated and implemented for human robots compared to the methods in which the parts of the search space are considered as the training data. The skipping of two persons' movement in opposite directions causes a mistake in identifying the group. It is even possible that two persons who are in the same group will not be together from the beginning and continue on the same path for sometimes, in this case it would be a mistake to distinguish the two persons from the proposed method at the beginning of the route.

a



b

Fig. 12. (a) Input video, (b) video after detecting social groups with the proposed method.

a

b

Fig. 13. (a) Input video, (b) video after detecting social group with the proposed method.

References

[1] Mehran, R., Oyama, A., Shah, M., : 'Abnormal crowd behavior detection using social force model,' in Computer Vision and Pattern Recognition, Conference on IEEE., 2009, pp. 935-942.

[2] Manzi, A., Fiorini, L., Limosani, R., Dario, P., Cavallo, F.: 'Two-person activity recognition using skeleton data', IET Computer Vision, vol. 12, no.1, pp. 27-35, 2017.

[3] Moussaïd, M., Perozo, N., Garnier, S., Helbing, D., Theraulaz G., :'The walking behaviour of pedestrian social groups and its impact on crowd dynamics,' PloS one, vol.5, no.4, pp. 10047-10054, 2010.

[4] Bandini, S., Gorrini, A., Manenti, L., Vizzari G.,: 'Crowd and pedestrian dynamics: Empirical investigation and simulation,' in Proceedings of Measuring Behavior, pp. 308-311, 2012.

[5] Van Stekelenburg, J., & Klandermans, B. Individuals in movements: A social psychology of contention. In Handbook of social movements across disciplines (pp. 103-139). Springer, Cham, 2017.

[6] Luber, M., Stork, J. A., Tipaldi, G. D. Arras, K. O.: 'People tracking with human motion predictions from social forces,' in Robotics and Automation (ICRA), 2010 IEEE International Conference on IEEE, pp. 464-469, 2010.

[7] Iqbal, T., Moosaei, M., Riek, L. D.,: 'Tempo adaptation and anticipation methods for human-robot teams,' in RSS, Planning HRI: Shared Autonomy Collab. Robot. Workshop, 2016.

[8] Iqbal, T., Rack, S., Riek, L. D.,: 'Movement coordination in human–robot teams: a dynamical systems approach,' IEEE Transactions on Robotics, vol.32, no.4, pp. 909-919, 2016.

[9] Vázquez, M., Steinfeld, A., Hudson, S. E.,: '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, pp. 3010-3017, 2015.

[10] Ricci, E., Varadarajan, J., Subramanian, R., Rota Bulo, S., Ahuja, N., Lanz, O.,: '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, pp. 4660-4668, 2015.

[11] Cristani M., et al.,: 'Social interaction discovery by statistical analysis of F-formations,' in BMVC, pp 4-11, 2011.

[12] Shao, J., Change Loy, C., Wang, X.: 'Scene-independent group profiling in crowd,' in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2219-2226, 2014.

[13] Pang, S. K., Li, J., Godsill, S. J.,: 'Detection and tracking of coordinated groups,' IEEE Transactions on Aerospace and Electronic Systems, vol. 47, no.1, pp. 472-502, 2011.

[14] Rodriguez, M., Laptev, I., Sivic, J., Audibert, J.-Y.,: 'Density-aware person detection and tracking in crowds,' in Computer Vision (ICCV), 2011 IEEE International Conference on IEEE, pp. 2423-2430, 2011.

[15] Yamaguchi, K., Berg, A. C., Ortiz, L. E., Berg, T. L.,: 'Who are you with and where are you going?,' in Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on IEEE, pp. 1345-1352, 2011.

[16] Chang, M.-C., Krahnstoever, N., Ge, W.,: 'Probabilistic group-level motion analysis and scenario recognition,' in Computer Vision (ICCV), 2011 IEEE International Conference on IEEE, pp. 747-754, 2011.

[17] Zanotto, M., Bazzani, L., Cristani, M., Murino, V.,: 'Online bayesian nonparametrics for group detection,' in Proc. of BMVC, 2012.

[18] Ge, W., Collins, R. T., Ruback, R. B.,: 'Vision-based analysis of small groups in pedestrian crowds,' IEEE transactions on pattern analysis and machine intelligence, vol.34, no.5, pp. 1003-1016, 2012.

[19] Park H. S., Shi, J.,: 'Social saliency prediction,' in Computer Vision and Pattern Recognition (CVPR), Conference on IEEE, pp. 4777-4785, 2015.

[20] Tran, K. N., Bedagkar-Gala, A., Kakadiaris, I. A., Shah, S. K.,: 'Social Cues in Group Formation and Local Interactions for Collective Activity Analysis,' in VISAPP, vol. 1, pp. 539-548, 2013.

[21] Shao, J., Dong, N., Zhao, Q.,: 'An adaptive clustering approach for group detection in the crowd,' in Systems, Signals and Image Processing (IWSSIP), 2015 International Conference on IEEE, pp. 77-80, 2015.

[22] Voon, W. P., Mustapha, N., Affendey, L. S., Khalid, F.,: 'A new clustering approach for group detection in scene-independent dense crowds,' in Computer and Information Sciences (ICCOINS), 2016 3rd International Conference on IEEE, pp. 414-417, 2016.

[23] Khan, S. D., Vizzari, G., Bandini, S., Basalamah, S.,: 'Detection of social groups in pedestrian crowds using computer vision,' in International Conference on Advanced Concepts for Intelligent Vision Systems, Springer, pp. 249-260, 2015.

[24] Yight, A., Temizel, A.,: 'Individual and group tracking with the evaluation of social interactions,' IET Computer Vision, vol. 11, no.3, pp. 255-263, 2016.

[25] Solera, F., Calderara, S., Cucchiara, R.,: 'Social constrained structural learning for groups detection in crowd,' IEEE transactions on pattern analysis and machine intelligence, vol.38, no.5, pp. 995-1008, 2016.

[26] Fernando, T., Denman, S., Sridharan, S., & Fookes, C., 'GD-GAN: Generative Adversarial Networks for Trajectory Prediction and Group Detection in Crowds,' In *Asian* Conference on Computer Vision, Springer, Cham, pp. 314-330, 2018.

[27] Reicher, S. D., Spears, R., Postmes, T.,: 'A social identity model of deindividuation phenomena,' European review of social psychology, vol.6, no.1, pp. 161-198, 1995.

[28] Eberhart R., Kennedy, J.: 'A new optimizer using particle swarm theory,' in Micro Machine and Human Science, IEEE, pp. 39-43, 1995.

[29] Peng, Hong, et al. "An automatic clustering algorithm inspired by membrane computing." Pattern Recognition Letters 68 pp. 34-40, 2015.

[30] Farsi, H., Mozaffarian MA., Rahmani, H.,: 'Improving voice activity detection used in ITU-T G.729.B,' Proc. of 3rd WSEAS International Conference on Circuits, Systems, and Telecommunications, pp. 11-15, 2009.

[31] Farsi, H.,: 'Improvement of minimum traching in minimum statistics noise estimation method,' Signal Processing An International Journal (SPIJ), vol. 4, no.1, pp. 1-17, 2010.

[32] Khosravi, H., Moradi, E., Darabi, H.,: 'Identification Of Homogeneous Groundwater Quality Regions Using Factor And Cluster Analysis; A Case Study Ghir Plain Of Fars Province,' 2015.

[33] Hosseini, S. M., Farsi, H., Yazdi, H. S.,: 'Best clustering around the color images,' International Journal of Computer and Electrical Engineering, vol.1, no.1, pp. 20-29, 2009.

[34] Hosseini, S. M., Nasrabadi, A., Nouri, P., Farsi, H.,: 'A novel human gait recognition system,' International Journal of Computer and Electrical Engineering, vol.2, no.6, pp. 1043-1055, 2010.

[35] Sadeghian, A., Kosaraju, V., Sadeghian, A., Hirose, N., Rezatofighi, H., & Savarese, S. Sophie,: 'An attentive gan for predicting paths compliant to social and physical constraints,' In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1349-1358, 2019.

[36] Wang, Q., Chen, M., Nie, F., & Li, X., 'Detecting coherent groups in crowd scenes by multiview clustering,' IEEE transactions on pattern analysis and machine intelligence, vol.40, no.7, pp. 910-919, 2018.

Ali Akbari received the B.Sc. and M.Sc degrees in Electronics engineering from University of Birjand, Birjand, Iran, in 2013 and 2015, respectively. Since 2015 he was accepted as a Ph.D candidate student in communications engineering, university of Birjand, Birjand, Iran. His area research interests include Image and Video Processing, Retrieval, Pattern recognition, Digital Signal Processing.

Hassan Farsi received the B.Sc. and M.Sc. degrees from Sharif University of Technology, Tehran, Iran, in 1992 and 1995, respectively. Since 2000, he started his Ph.D in the Centre of Communications Systems Research (CCSR), University of Surrey, Guildford, UK, and received the Ph.D degree in 2004. He is interested in speech, image and video processing on wirelesscommunications. Now, he works as professor in communication engineering in department of Electrical and Computer Eng., University of Birjand, Birjand, IRAN.

Sajad Mohamadzadeh received the B.Sc. degree in communication engineering from Sistan & Baloochestan, University of Zahedan, Iran, in 2010. He received the M.Sc. and Ph.D. degree in communication engineering from South of Khorasan, University of Birjand, Birjand, Iran, in 2012 and 2016, respectively. Now, he works as assistant professor in Faculty of Technical and Engineering of Ferdows, University of Birjand, Birjand, Iran. His area research interests include Image and Video Processing, Retrieval, Pattern recognition, Digital Signal Processing, Sparse Representation, and Deep Learning.