

This document is published in:

Molina, J. M.; et al. (eds.) (2011) *User-Centric Technologies and Applications: Proceedings of the CONTEXTS 2011 Workshop*, (Advances in Intelligent and Soft Computing, 94) Springer, pp. 47- 57.
Doi: http://dx.doi.org/10.1007/978-3-642-19908-0_6

© 2011 Springer-Verlag Berlin Heidelberg

A structured representation to the group behavior recognition issue

Alberto Pozo¹, Jesús Gracia¹, Miguel A. Patricio¹ and José M. Molina¹,

¹ GIAA, Carlos III University, Spain

{ alberto.pozo, jesus.garcia, miguelangel.patricio, josemanuel.molina }@uc3m.es

Abstract. The behavior recognition is one of the most prolific lines of research in recent decades in the field of computer vision. Within this field, the majority of researches have focused on the recognition of the activities carried out by a single individual, however this paper deals with the problem of recognizing the behavior of a group of individuals, in which relations between the component elements are of great importance. For this purpose it is exposed a new representation that concentrates all necessary information concerning relations two to two present in the group, and the semantics of the different groups formed by individuals and training (or structure) of each one of them. The work is presented with the dataset created in CVBASE06 dealing the European handball

Keywords: Group behavior recognition, activity representation, computer vision.

1 Introduction

Human activity analysis and behavior recognition has received an enormous attention in the last two decades of computer vision community. A significant amount of research has addressed to behavior recognition of one element in the scene.

Instead of modeling the activities of one single element, group behavior recognition deals with multiple objects and/or people, who are part of groups.

Group behavior recognition represents a relatively new interesting direction of research, which has many possible applications in different situations like group sports, surveillance, defense, ethology, etc.

In behavior recognition there are two distinct philosophies for modeling a group; the group could be dealt as a single group (crowd) or as a composition of individuals with some shared objectives. In this paper we focus the investigation in the second philosophy, where take place many distinguishable agents.

Individual behavior recognition and group behavior recognition have differences we should consider. Group behavior is not the addition of multiples individual's behaviors, instead of, group behavior depends on individual's activities, relations between the elements and roles played by each element.

The recognition of the dynamic of the groups can be applied in many complex areas, such as sports, security, ethology and defense. This recognition also can be

extrapolated to any field of research, composed by autonomy agents which behaviour needs to be studied.

The present paper shows a new representation of the possible variables existed in the problem. This had been designed to put in order briefly the essential information of the system.

With the aim of achieve our project, it will rely on three levels of abstraction.

Firstly, a matrix will be established to store away the polar coordinates position at the system and the binary relationship existed between them. They will be represented by the free vectors which connect both individual.

In these terms and conditions, for each frame in the video, the geometrical information has being kept in two or three dimensions.

Once being contained all the geometrical information, the process continues in a second abstraction's level where the challenge is capturing the logical information implicated between the communication of individual and groups. For this reason it is necessary to make different combinations for representing every group of the system.

It is a relevant detail to remark that each individual can belong to a several groups at the same time, and the groups have the possibility to incorporate an undefined number of other groups or individual.

In the third level, a new representation is created to reduce the dimension of the problem. One of the important key in this type of domains is that the number of relations between the elements of the scene growth exponential in relation with the number of elements. For this reason, a new representation is created to save the formation information of each group without saving all the relations between each element. Instead of save all the possible edges in a graph, this approach only save important graph that can provide all the formation information wasting less space.

The paper is organized as follows. Section 2 reviews related work. Section 3 describes the problem. Section 4 introduces our description. Conclusions are drawn in section 5.

2 Related work

Despite the fact that there is plenty of work on single object activities (1), (2), (3), the field of group activities recognition is relatively unexplored.

Group behavior recognition is a complex task that can be observed under different points of view.

There are two big families of approaches, one logical and one geometrical.

The logical approaches (4) are focused in construct a context-free grammar to describe the group activity based in the individual activities of the group's member.

The main characteristic of this point of view is the important of the first level, the features extraction. They need a previous system that recognizes the activity of each element of the scene.

The geometrical approaches (5), (6) have a different point of view. The features extracted in this case are based on the coordinates of the elements of the scene.

This approaches use to have higher computational complexity and the number of the elements in the scene could become very important.

There are also approaches than combine both perspective, like (7) whose work recognize key elements of a sport (basketball) using the positions of the players. This approach needs to identify the key elements of the domain dealt, and these key elements could be different in many different situations.

One more general approach could be read in (8) where the trajectories of the players (in a Robocup mach) are coded to create set of patterns that identify each type of action.

3 Group behavior recognition issue: a general overview

Group behavior recognition is composed by two steps: in the first one the features of the system should be extracted, and in the second one the features are used to recognize the behavior.

In this paper we are going to focus in the second step, we use the extracted features by other system to make the inference of the behavior.

3.1 General Description

In a general scene exist one area composed by many sub-areas and a number of elements (which could be fixed or not).

Each element of the system has a set of features (like positioning, color, shape, etc.) The element's features could suffer changes in time.

Each element of the system should belong to a group, and could belong to many groups at the same time.

It is important to emphasize that any element of the system must be in a group, so there are not isolated elements.

Each group has a internal and a external attitude. Each attitude could be cooperative or competitive. Internal attitude defines the attitude between the members of a group, and external attitude defines the attitude of the group respecting of the rest of the groups.

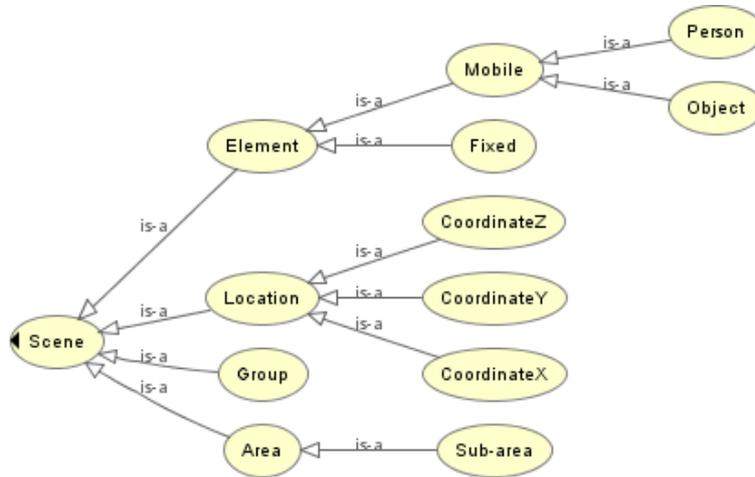


Fig. 1 General scene

Fig. 1 shows the general scene described above.

3.2 Problem description

Some of the general axioms of the problem describe above have been eliminated for more practical approach of the problem.

In our approach there is one sequence of video composed by a number of T frames, where are included a number of N elements (this number cannot change in time). The elements of the scene are distributed in a number of G groups, and each group is represented by a graph. (Group could be composed by two or more elements, and one element could be part of one or more groups).

Each node constitute an agent of the group, and each graph is represented by his edges

One edge is represented by free vector expressed in polar coordinate system.

Fig. 2 shows a scene with six elements conforming three groups. The definition of the groups is the semantic representation of the relations between the elements of the system.

The features selected to describe the elements of the scene are its coordinates (in polar coordinate system or spherical coordinate system in case of 3-D positioning) and the coordinates of the free vectors that represents the edges of the graphs.



Fig. 2 Graphic representation of a system with six elements and three groups

For each element and each possible edge we save the coordinates of the vector for each frame of the scene. One position vector for each element and M free vectors for the edges, where $M = \frac{N*(N-1)}{2}$.

To describe the spatial relation between elements i and j in the frame t , there is a vector called $v_{i,j,t}$ where $r_{i,j,t} = |r_{i,j,t}|^{\theta_{i,j,t}}$, with $|r_{i,j,t}| = \sqrt{(x_{j,t} - x_{i,t})^2 + (y_{j,t} - y_{i,t})^2}$ and $\theta_{i,j,t} = \tan^{-1} \frac{(y_{j,t} - y_{i,t})}{(x_{j,t} - x_{i,t})}$.

4 A structured representation to the group behavior recognition issue.

In most the case of group behavior recognition it is difficult to identify any individual action and role played by each element of the group. Otherwise, tracking a group of elements in computer video analysis could provide the positioning of the elements in the scene.

There are a lot of systems providing the location of elements of a group, like GPS (outdoor), UWB (indoor), tracking, etc.

Behavior recognition based on the positioning of each element of the group could provide results without knowing the activity carried out by the elements of the group.

We propose a structured representation composed by three matrix called R , A and S . The first one save the geometrical data of the elements in time (the position vectors and free vectors), while the second one represents the information about the semantic of the scene, composed by the number of groups founded and their makeup, and the third one represents the shape of each graph of the scene.

This structured representation contains the information about the position of each element of the scene, the information about the spatial relations between the elements, and the groups shape information. It is importance to notice that the last one information could be representative of the behavior of the group in many different domains, like group sports, surveillance, defense, ethology, etc.

4.1 Geometrical information

All the geometrical information about the elements in the scene, and its relations is saved on matrix R, this information is used to construct the shape matrix.

Matrix R is a three dimensional matrix with the information of the location of each agent and the free vector that represent each possible edge presented at the scene.

A scene with N elements has $= \frac{N*(N-1)}{2}$ possible edges that must be saved, and N position vectors.

Each vector of the matrix has two or three components (it is depending on at the cases a 3-D scene or 2-D scene).

The first ones N vectors represent the location of the N elements, and the next M vectors represent the edge of the graphs.

The R matrix has one row for each frame of the scene and N + M columns.

$$\begin{pmatrix} |r_1|^{\theta_1} & \dots & |r_N|^{\theta_N} & |r_{1,2}|^{\theta_{1,2}} & \dots & |r_{N-1,N}|^{\theta_{N-1,N}} \\ \vdots & \ddots & & & \ddots & \\ |r_1|^{\theta_1} & \dots & |r_N|^{\theta_N} & |r_{1,2}|^{\theta_{1,2}} & \dots & |r_{N-1,N}|^{\theta_{N-1,N}} \end{pmatrix}$$

4.2 Semantics information

The semantics information represents the associations between the elements of the scene to perform groups. One element could be part of many groups, and could be many groups.

This information makes it possible to create different associations between elements to grasp better the semantics context.

This semantics information is saved in a binary matrix with one row for each group, and one column for each element.

The matrix can only contain zeros or ones, which represent if this element forms part of the graph.

For example, in a scene with nine elements, and two groups, the matrix A could be like this: $A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$. This matrix shows that there are two graphs, the first one composed by the elements: 1, 2, 3, 4, and 5; and the second one composed by the elements: 5, 6, 7, 8 and 9.

4.3 Shape information

Matrices S define the shape of the graphs, there are one matrix for each graph (row of the A matrix).

Each S matrix has a number of T rows, and one column for each possible edge of the graph. The number of possible edge depends on the number of elements that composed the graph ($M_g = \frac{N_g*(N_g-1)}{2}$).

Elements of S have two components: relative distance (d) and direction (γ).

Relative distance is calculated by the formula $d = \left\lceil \frac{|r_{ij}| * 8}{|r_{max}|} \right\rceil$. Where r_{ij} is the distance between the elements i and j , and r_{max} is the maximum distance between any elements of the graph. By definition d is a natural number between 1 and 8.

Direction between two elements of the graph is defined by the formula $\gamma = \left\lceil \frac{|\theta_{ij}| * 8}{\pi} \right\rceil$. Where θ_{ij} is the angle between the elements i and j , and. By definition γ is a natural number between 1 and 8. It is important to perceive that in spite of the graphs are not directed, to construct the reduced graph we have to distinguish between the same directions with different sense. So the possible directions are covered between $-\pi$ and π radians.

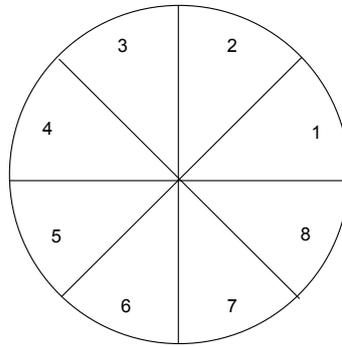


Fig. 3 Directions code

Each S matrix has the edges of the graph that defines its shape. If some edge of the element i have the same direction that another one and it is longer than the previous one, then this edge is not added to the matrix. One negative value (-1) is added in this position.

Figure 4 shows the construction process: the shortest edge of the first element is added in (a). Then the second shortest is also added in (b). In (c) there is a shorter edge with the same direction (2) and the edge is not added. The process is repeated until all the elements are checked (d), (e) and (f).

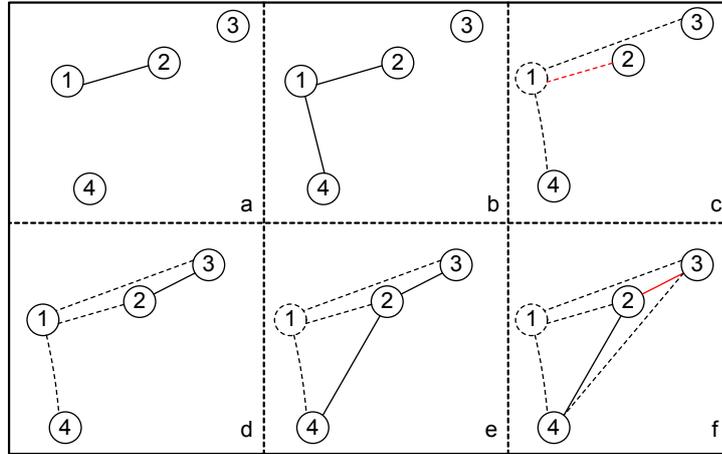


Fig. 4 Construction process

The matrix below shows the S matrix of the graph in the Fig. 4. First row represents the graph at the instant $t = 0$, and last row shows the graph at the instant $t = T$.

Fig. 4 shows an example of a graph with five elements. In the first frame the edges between the nodes 1-5, 3-4 and 3-5 are not defined because they have the same directions (and they are longer) than the edges 1-5, 3-2 and 5-2.

Then, in the frame T the graph's shape has changed, there are new relevant edges (like 3-5 and 3-4) and some relative distance have also changed.

$$\begin{pmatrix} 4^6 & 6^8 & 3^4 & -1^5 & 3^2 & 4^1 & 5^3 & -1^1 & -1^3 & 4^5 \\ \vdots & \vdots \\ 6^6 & 5^8 & 4^4 & -1^5 & 4^5 & 6^8 & 5^2 & 6^1 & 7^3 & 5^5 \end{pmatrix}$$

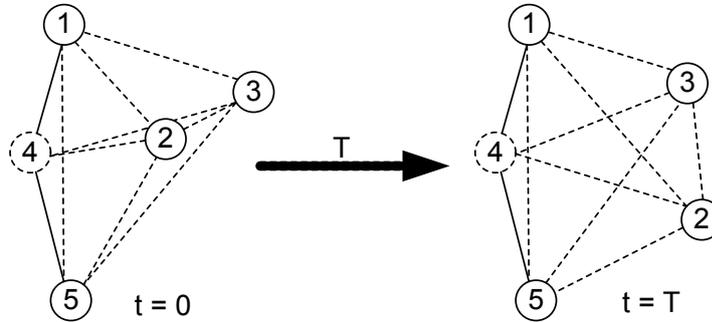


Fig. 5 Example graph

5 CVBASE06 dataset

In the previous section, we define a structured representation for group behavior recognition. These paragraphs are going to show a practice application for the European handball domain.

We have used a dataset called CVBASE06 dataset, from: <http://vision.fe.uni-lj.si/cvbase06/downloads.html>.

CVBASE06 dataset (9) has three sports: squash, basketball and European handball. We have selected the handball dataset because it has the positioning of the player and the action played by the team at each moment.

5.1 European handball CVBASE06 dataset

The handball CVBASE06 dataset part has information about the players (their position) and about the teams (the action played), distributed in six files.

There are ten minutes of video recorded with three cameras (A, B and C), all of them synchronized.

The video has 25 fps, so there are 15000 frames. Each frame has the position of each player of a team, and the group action played by this team.

Each group action start when finish the previous one, so in each frame there is one, and only one, group action.

A European handball team is composed by seven players, one of them is a special player called goalkeeper.

5.2 Active attack and passive attack

There are two different types of attack in European handball, one of them aims to score quickly, and the other one tries to keep the ball. The first one is called an active attack, and the second one is called passive attack.

The two different types of attack are labeling with: “nfan” and “nfpn” respectively.

For this domain we have define the A matrix to create a group with four players of the attacking team.

The figure 6 shows the S matrix for two sequences of the video, in the first one (a) we can see a passive attack, and in the second one we can see a active attack (b).

The “frame” axis represents the evolution of time, in frames. The “edge” axis represents each edge of the graph, numbered on (c). The (d) shows a frame of the scene, with the elements of the graph colored.

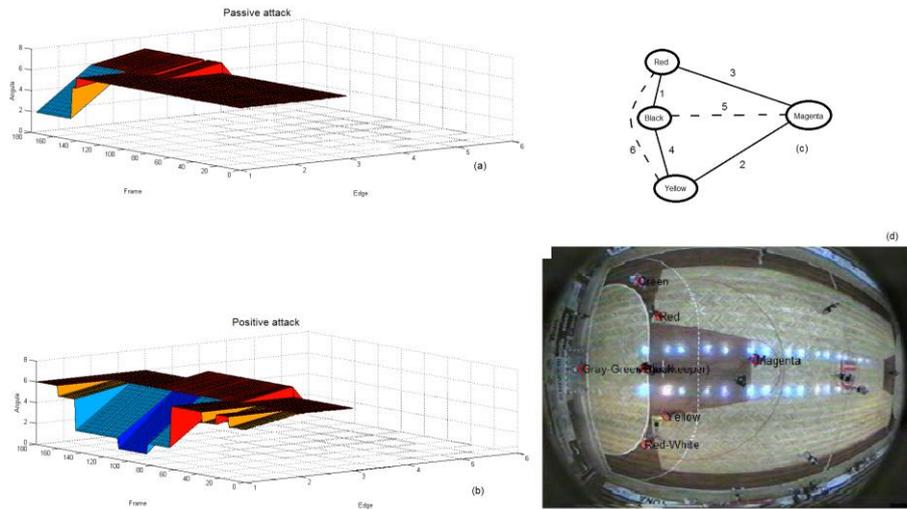


Fig. 6 Passive/Active attack

The different dynamism of the two types of attack is reflected on the “frame” axis. In the first attack, there are few changes, but in the second one there are a lot of changes.

In the fig. 6 (a) and (b) also we can see that the graph constructed with the S matrix. In this graph the algorithm has deleted the edges 5 and 6, edges 1, 2, 3 and 4 have all the information about the formation of the group.

6 Conclusions

There are a lot of papers related with the behavior recognition, but most of them are related with one single element. This paper had focused on the group behavior recognition, where a group of elements compose a group and have a group behavior.

In the group behavior recognition the researches have followed two ways approach: logical and geometrical. Because there are a lot of systems providing the location and tracking of elements of a group, this paper had focused on the geometrical approach.

This approach had reduced the number of relations without loss information about the formation to realize the rezoning process. This approach is based in a novel structured representation of the important relations between the elements of the graphs.

This structured representation save the information about the formation of the group, wasting less space.

One of the assumption makes is that the formation information of the group (and its dynamics) is representative of the group behavior. This assumption could be applied to many different domains like: defense, etholy, group sports, etc.

Acknowledgements: This work was supported in part by Projects CICYT TIN2008-06742-C02-02/TSI, CICYT TEC2008-06732-C02-02/TEC, CAM CONTEXTS (S2009/ TIC-1485) and DPS2008-07029-C02-02.

References

1. **Aggarwal, J and Cai, Q.** Human motion analysis: a review. *Computer vision and image understanding*. pp. 428-440.
2. **Bobick, A and Wilson, A.** A state-based approach to the representation and recognition of gesture. *PAMI*. pp. 1325-1337.
3. **Moeslund, T B, Kruger, V and Hilton, A.** A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*. 2006, pp. 90-126.
4. *Recognition of High-level Group Activities Based on Activities of Individual Members.* **Ryoo, M. S. and Aggarwal, J. K.** 2008, WMVC.
5. *Detecting Group Activities using Rigidity of Formation.* **Khan, Saad M. and Shah, Mubarak.** Singapore : ACM, 2005.
6. *Learning Multi-modal Densities on Discriminative Temporal Interaction Manifold for Group Activity Recognition.* **Ruonan, Li, Rama, Chellappa and Shaohua, Kevin Zhou.** New York : s.n., 2009.
7. *A trajectory-based analysis of coordinated team activity in a basketball game.* **Perse, Matej, et al.** Ljubljana : s.n., 2009.
8. **Ramos, Fernando and Ayanegui, Huberto.** Tracking behaviours of cooperative robots within multi-agent domains. Tlaxcala : s.n.
9. **Pers, Janez.** *CVBASE 06 Dataset: A Dataset for Development and Testing of Computer Vision Based Methods in Sport Environments.* Ljubljana : s.n., 2005.
10. *Representation and visual recognition of complex multi-agent actions using Belief networks.* **Intille, S and Bobick, A.** Freiburg, Germany : s.n., 1998. ECCV Workshop on Perception of Human Action.