Research Article

A Feature Selection Approach to the Group Behavior Recognition Issue Using Static Context Information

Alberto Pozo, Miguel A. Patricio, Jesús García, and José Manuel Molina

University Carlos III of Madrid, 28270 Colmenarejo, Spain

Correspondence should be addressed to Alberto Pozo; alberto@pozoesteban.es

This paper deals with the problem of group behavior recognition. Our approach is to merge all the possible features of group behavior (individuals, groups, relationships between individuals, relationships between groups, etc.) with static context information relating to particular domains. All this information is represented as a set of features by classification algorithms. This is a very high-dimensional problem, with which classification algorithms are unable cope. For this reason, this paper also presents four feature selection alternatives: two wrappers and two filters. We present and compare the results of each method in the basketball domain.

1. Introduction

Human or other types of behavior recognition are currently one of the most prolific fields of research. One typical restriction in this field is that the scene should contain only one element. The system analyzes the behavior of the element and recognizes what the element is doing. Several papers address this aim [1, 2]. However, many situations, like team sports, animal social behavior (ethology), traffic analysis, and so forth, hinge on more than one element, where elements behave socially with respect to the other elements of the group. Individual behavior depends on group behavior in such cases. Therefore, individuals have to be studied together, as a group. Research on this area is scant. Some researchers deal with the elements as a crowd and try to recognize their behavior by analyzing shape or some other features. On the other hand, some authors deal with each element as an individual that is a member of the group. Our proposal takes this approach. Behavior recognition can be divided into two steps: feature extraction and feature analysis. In feature extraction, the system chooses and extracts the main raw information (from video cameras, GPS, RFID, TOF cameras, etc.) and builds a set of features. Features could be raw information, like location, or derivatives, like velocity (from location in time). Features may or may not be context dependent. During feature analysis, the system transforms this raw information and recognizes the underlying behavior. In most cases, feature extraction is a distributed problem, since device networks (e.g., video cameras) can provide information on the scene from different viewpoints. In our particular case, the domain is inherently distributed because we have a four-camera network providing information on a complex scene in which single viewpoints are frequently occluded. The research reported in this paper is part of a larger project, aiming to discover high-level information (recognition of group activities) by extracting the features of the scene using a distributed network of devices. We propose a new representation of the selected features that could be used to learn and predict group behavior. Although a scene may, in some domains, have a great many features and feature types may vary, we believe that the identification of the group elements is likely to be very meaningful in most potential application domains. Such problems could have a wide range of possibilities. For this reason, static context information is essential for our purpose. For instance, we need to know how many groups there are, how many elements the groups contain, and whether the number of groups or of group members is static or dynamic. All this static context information depends on the specific domain. Also, some domains could have a dynamic context that could be incorporated to our representation to improve the accuracy of the classification. In the basketball domain,

for example, remaining match time or score could provide useful information. Finally, these features are structured at different levels depending on whether they relate to an individual, the scene, or the group. Our proposed representation is divided into three different levels, which are described below. Some of the selected features are domain specific, but features related to the location of the elements could be generalized to most domains. To prove the worth of the representation, we have classified the dataset group behavior [3]. To do this, we used a well-known technique called hidden Markov model [4]. This representation could be useful for the group behavior recognition issue. Also, we selected the most representative features for the specific domain proposed in [3]. There are a great many potentially useful general features in the basketball domain, like player position, who is the ball handler, and so forth. But other features could be created by processing the raw data and creating context-dependent features. For instance, if there is information about which player has the ball, this data could be used to create location features like players relative position with regard to the ball handler. In this case, we have static context information about the number of groups (two) and the number of elements in each group (two). The paper is organized as follows. Section 2 reviews related work. Section 3 describes the problem issue. Section 4 reports the experiments. Conclusions and future work are outlined in Sections 6 and 7.

2. Related Work

Behavior recognition is a very active research line in the computer vision field. Many researchers have tried to extract high-level information about human behavior from different types of sensors, especially from video [1, 2, 5], and also from other sensor types, like accelerometers, gyroscopes, and so forth. All these approaches focus on individual human behaviors, but humans are social beings, and a lot of behaviors are dependent on the other elements in the group. For this reason, there is a relatively new area of research focusing on elements as members of a group. The behavior of each element in a group depends on the behavior of the other elements, and the combination of individual behaviors results in a high-level group behavior. This type of behavior is observed in many situations, like RoboCup [6], team sports [7–9], a military parade [10], automatic camera surveillance in public places [6, 11, 12], and so forth. Depending on the type of analyzed features, all these works could be classified into two main streams: logical reasoning and geometrical reasoning. Logical approaches, like [12, 13], are based on preprocessed high-level features, where, for example, player trajectories are discretized into major subareas, and some key movements, like block, are detected [8]. This high-level information is then analyzed to recognize team behavior. Such approaches are typically very effective, but, depending on high-level features extraction system, these solutions tend to be too ad hoc for the system to be generalized. On the other hand, there are geometrical approaches that use lower-level features, like raw position, or a derivative feature, like speed or acceleration. For example, player trajectories in an American football game are used in [13] to generate a discriminative

temporal interaction matrix, using a 4D tensor and a refactor 2-tensor kernel. This approach is designed to classify five different team behaviors. All these approaches focus on the second step of behavior recognition, where the features have already been extracted, and the system uses these features to classify or predict group behavior. So, they need a first step that extracts the selected features from the raw data, for instance, extracts American football player trajectories from the match video data. Another possibility is to use an existing dataset to develop, improve, and test the approach. But there are not many suitable datasets in group behavior recognition, and this is a very important obstacle. For example, papers like [13] use a nonpublic dataset called GaTech Football Play, and there are other commercial approaches, like Prozone, where eight fixed cameras are positioned on a soccer field to capture the trajectory of all players in the match. Unfortunately, we do not have access to this information. On the other hand, the Namez Pers work [14] is divided into three different parts related to three different sports: squash, basketball, and European handball. The European handball team sport section has the necessary information for the group behavior recognition research field. This information is composed of one team player positions and the team behavior in each team for a ten-minute video. Like [14], there is information on player position in a 2-on-2 basketball situation in [3], including information regarding the ball handler and the type of attack (from eight different types). This dataset [3] was used in this paper to develop and test our approach. For the final classification and prediction, we have used a well-known technique called hidden Markov models [4]. The dimensionality of problems in group behavior recognition domains is usually very high, because they include more than one element, each of which could have a different set of features possibly resulting from the relationship between elements. For this reason, problem dimensionality reduction through feature subset selection is potentially very useful for the group behavior recognition issue. Dimensionality could be reduced using wrappers or filters. Wrapper methods use a predictive model to score feature subsets. Each new subset is used to train a model, which is tested on a holdout set. Counting the number of mistakes made on that holdout set (the model error rate) gives the score for that subset. As wrapper methods train a new model for each subset, they are very computationally intensive but usually provide the best-performing feature set for that particular type of model. Many papers, like [15] or [16], use this kind of algorithms. Filter methods use a proxy measure instead of the error rate to score a feature subset. This measure is designed to be fast to compute while still capturing the fitness of the feature set. Common measures include mutual information [17–19], the Pearson product-moment correlation coefficient [20], and the inter-/intraclass distance [21]. Filters are usually less computationally intensive than wrappers, but they produce a feature set which is not tuned to a specific type of predictive model.

3. Group Behavior Recognition Issue

Group behavior recognition is a new field of research resulting from the elimination of the one-element restriction in

the activity recognition issue. This field of research has a lot of potential application domains, such as team sports, military intelligence, fauna behavior recognition, and video surveillance. A characteristic of the behavior recognition research field is that action happens in time. A sequence of observations rather than a single observation has to be classified, and each observation depends on the previous one. This characteristic results in two problems: classification and segmentation. In classification, there are several segmented sequences, where each segment contains a single group behavior; in segmentation, there are raw sequences with different group behaviors, which have to be segmented. This paper focuses on the first approach, where we have several segmented sequences. The classification phase of the group behavior recognition is composed of two steps: extraction and recognition. During extraction, the raw data must be extracted from the system to identify features. During recognition, these features are selected and used to recognize the behavior. The system could have a lot of feature types, like position, individual action, trajectory, speed, color, and so forth. In this paper, we focus on recognition. After a short description of the dataset construction process (feature extraction), we illustrate how this information could be used for behavior recognition.

3.1. A General Problem Description.

Group behavior recognition could be applied in a range of domains, as described above. All these domains have some common features described below. There is a scene S, composed of a number M of groups, a number F of features, and dynamic context information C (1). The scene includes group behaviors from a set of behaviors b_p called B (2). All features and behaviors are time dependent:

$$S_t = \left\{ G_t^1, G_t^2, \ldots, G_t^M, F_t^1, F_t^2, \ldots, F_t^{O_S}, C \right\} \mid 1 \le t \le T, \quad (1)$$

$$B = \left\{ b_1, b_2, \ldots, B_p \right\}. \quad (2)$$

Each group is composed of a number of domain-dependent elements N_m and a number of domain-dependent features O_G (3). Also, each group performs a group behavior from B (2):

$$G_t = \left\{ I_t^1, I_t^2, \ldots, I_t^{N_m}, H_t^1, H_t^2, \ldots, H_t^{O_G} \right\} \mid 1 \le t \le T. \quad (3)$$

Finally, each element is composed of a number O_I of domain-dependent features. All these features are changing, and the element n is defined at time t by

$$I_t^n = \left\{ J_t^1, J_t^2, \ldots, J_t^{O_I} \right\} \mid 1 \le t \le T. \quad (4)$$

Hence, the scene has M groups and O_S features, where each group is composed of N_M elements and O_G features, and each element has O_I features. There are four feature types: positive, Boolean, relative, and enum. Each type describes an element feature.

(i) Positive, described by (5), represents a feature that could be any value greater than zero. This could typically represent the positioning on one axis, a distance, and so forth:

$$\left\{ p_{nf} \mid f = 1 \cdots F, n = 1 \cdots N, p = 0 \cdots \inf \right\}. \quad (5)$$

(ii) Boolean, described by (6), could be any binary feature that represents whether or not the element has a characteristic, for instance, whether a player has the ball:

$$\left\{ b_{nf} \mid f = 1 \cdots F, n = 1 \cdots N, b = \{0, 1\} \right\}. \quad (6)$$

(iii) Relative, described by (7), represents a feature whose value must be between a minimum and a maximum value, for example, sensor luminosity:

$$\left\{ r_{nf} \mid f = 1 \cdots F, n = 1 \cdots N, r = \{0, 100\} \right\}. \quad (7)$$

(iv) Enum, described by (8), could be useful if the elements have a characteristic taken from a list of values, for instance, the role played by a basketball player:

$$\left\{ e_{nf} \mid f = 1 \cdots F, n = 1 \cdots N, e = \{1, 2, 3, \ldots, K\} \right\} \quad (8)$$

$$K = \text{"possible values of feature"}.$$

Exactly which features to select for extraction is a very important decision that usually depends on the problem domain. However, there are some features that are noteworthy in most potential domains; for instance, the location of each group element is almost always related to group behavior. Our approach is able to represent the entire group behavior recognition scene for both classification and segmentation. In this paper, we have focused on classification, where there is not only one sequence with several behaviors but also several sequences with a single behavior. In the learning phase, the system has lot of sequences S_1, S_2, and so forth, each is composed of a number of times T_1, T_2, and so forth and labeled with only one behavior b_p from the set B (2). We also analyze a new sequence S_j and predict its label b_j in the prediction phase.

3.2. Defending Pick-and-Roll Move Problem Domain.

This paper focuses on the 2-on-2 basketball domain. This is a very fitting problem because it represents a complex domain, like a basketball match, with just a few (only four) elements. In team sports, group behavior is often performed by only a subgroup of the team. In soccer, for example, the defense line is composed of only four or five team members, and that subgroup performs a lot of analyzable group behaviors. Also, in basketball matches, just two players from the same team could be involved in the pick-and-roll move. This is the move that the defenders try to stop in [3]. There is a sequence segmented into 23 different moves, where each segment represents an instance of five types of defense move. These segments picture four players: two attackers and two defenders. The two attackers make several attempts at the same move (pick and roll), and the two defenders try to

stop them in five different ways. These defensive moves are the group behavior to be predicted and are labeled fight/go over, go below, show and recover, push, and trap. The dataset used provides information gathered by four cameras at four different locations around the scene. Information from different viewpoints is critical in environments where occlusions are common. This information is obtained by tracking the players. Additionally, an expert (basketball coach) views this information in order to label tactical moves.

4. Features

The extracted features are the X and Y position and player handling the ball at any time. X and Y are represented by a positive feature type (see (5)) and the ball handler by the enum feature type (8). From this raw information we have built forty features, composed of thirty-nine positive features and one enum. These features describe the player positions (with eight features, four players, and two axes); team center positions (with four features, two teams, and two axes); all player center position (with two features); relative player positions with regard to ball handlers (with eight features); player velocity (with eight features, four players, two axes, and a ten-frame window); team velocity (with four features, two teams, two axes, and a ten-frame window), and all player center velocity (with two features). Finally, there are three features from the Laplacian invariant of the graph. Also an enum feature indicates who the ball handler is. All these features are summarized in Table 1, and some are shown in Figures 1, 2, 3, and 4.

The proposed representation has $M = 2$ groups, $N = 4$ elements, where each group has $N_1 = 2$ and $N_2 = 2$ elements, respectively, and the scene has five features $O_S = 8$ (the ball handler, the location and velocity of the center of all players, and the Laplacian invariant of the graph), each element has $O_I = 6$ features (XY position, XY velocity, and XY relative position), and each group has four features $O_G = 4$ (XY position and XY velocity). This dataset was used for all the experiments [3], but, in order to apply all the algorithms in the MATLAB framework, we have flattened the representation to two dimensions as described below. We have omitted the time dimension, and all other dimensions have been stored consecutively in the order described as follows:

$$
\begin{aligned}
S_t &= \left\{ I_t^1, I_t^2, I_t^3, I_t^4, G_t^1, G_t^2, F_t \right\} \quad \text{with} \\
I_t^n &= \left\{ X_t^n, Y_t^n, XX_t^n, YY_t^n, Vx_t^n, Vy_t^n \right\}, \\
G_t^m &= \left\{ X_t^m, Y_t^m, Vx_t^m, Vy_t^m \right\}, \\
F_t &= \left\{ X_t, Y_t, Vx_t, Vy_t, H_t, L_t^1, L_t^2, L_t^3 \right\},
\end{aligned}
\tag{9}
$$

where X and Y represents the coordinates in the couch system; XX and YY represent the coordinates of one element with regard to the ball handler; V_x and V_y represent the velocity; H indicates the ball handler, and L is the invariant Laplacian feature. In this case, we unfortunately have no dynamic context information that we could use to improve system accuracy.

TABLE 1: Features list.

Number	Feature
1	Player1's X position
2	Player1's Y position
3	Player2's X position
4	Player2's Y position
5	Player3's X position
6	Player3's Y position
7	Player4's X position
8	Player4's Y position
9	X coordinate of the center of the attacking team
10	Y coordinate of the center of the attacking team
11	X coordinate of the center of the defending team
12	Y coordinate of the center of the defending team
13	All players' X position
14	All players' Y position
15	Ball handler
16	Player1's X position with respect to the ball handler
17	Player1's Y position with respect to the ball handler
18	Player2's X position with respect to the ball handler
19	Player2's Y position with respect to the ball handler
20	Player3's X position with respect to the ball handler
21	Player3's Y position with respect to the ball handler
22	Player4's X position with respect to the ball handler
23	Player4's Y position with respect to the ball handler
24	Player1's X velocity
25	Player1's Y velocity
26	Player2's X velocity
27	Player2's Y velocity
28	Player3's X velocity
29	Player3's Y velocity
30	Player4's X velocity
31	Player4's Y velocity
32	X velocity of the attacking team
33	Y velocity of the attacking team
34	X velocity of the defending team
35	Y velocity of the defending team
36	All players' X velocity
37	All players' Y velocity
38	First component of the Laplacian invariant of the graph
39	Second component of the Laplacian invariant of the graph
40	Third component of the Laplacian invariant of the graph

We used a hidden Markov model algorithm for all experiments [4]. For the global classifiers, which classify the example into one of the five move classes, we have used three Gaussians per class configuration. For the binary classifiers, which classify whether or not the example belongs to a specific

Figure 1: Some features (1/4).
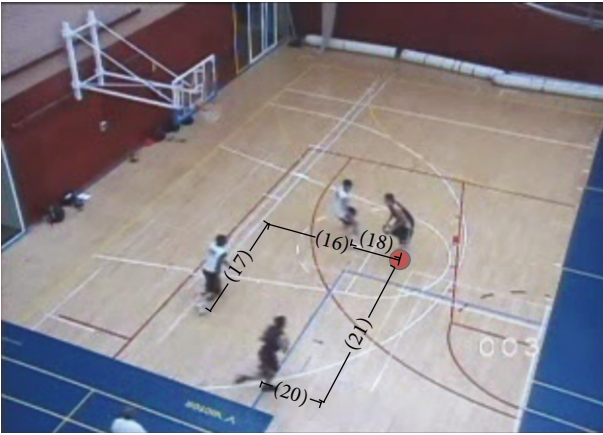

Figure 3: Some features (3/4).

Figure 2: Some features (2/4).

Figure 4: Some features (4/4).

class, we have used two Gaussians for member of the class and four Gaussians for not a member of the class. These configurations are constant across all experiments and feature selections.

As mentioned above, the dimensionality of group behavior recognition domains is usually high. In our basketball domain, we have forty features at any one time. So, it is very important to select a subset of features that would be able to provide all the necessary information for behavior recognition. In some cases, the information provided by a subset of features could be better than the information provided by all the features, because some features may, under some circumstances, only add noise.

4.1. Features Selection. In this paper we propose a comparison between four feature selection systems: global wrapper, binary wrapper, global mRMR filter, and binary mRMR filter.

4.2. Global Wrapper. From the features provided by the dataset [3], we have identified forty features as mentioned above. We have selected some of all these features following the algorithm described in [15]. This algorithm has a list of candidate features, composed in this case of the forty features

described above (see Table 1). The system then calculates the accuracy using only one of the features and selects the feature that has the best outcome. This feature is moved from the candidate list to the final list. However, the system calculates the accuracy with the selected feature and each one of the candidates features and again selects the feature that provides the best accuracy. The algorithm stops when the addition of another feature fails to improve accuracy. Figure 5 illustrates this process.

The diagram shows the algorithm in Figure 5. The OR logical gate symbol indicates that only one of the features is entered, and the AND logical gate symbol indicates that several features are combined in the classification. Figure 6 shows the system results with only one feature in the first step of the algorithm. The rows represent the sequences, and the columns represent experimentation with one feature. Red means that the system classified the example incorrectly, and green means that the system succeeded in classifying the sequence.

In this case, the selected feature was number 22, which represents player four X position with respect to the ball handler. At the end of the features selection process, five
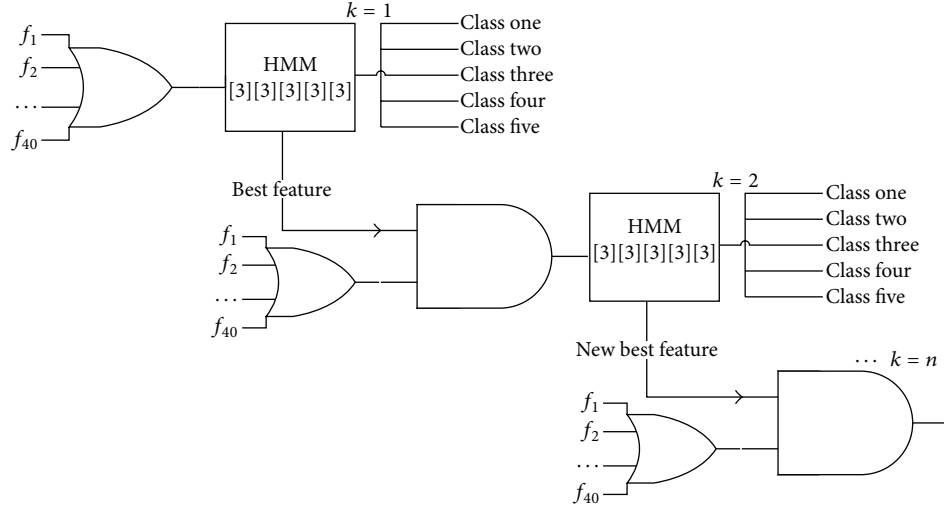
FIGURE 5: Global wrapper algorithm.

FIGURE 6: One feature matrix.

features were selected. Figure 7 shows the order of selected features and system accuracy.

4.3. Binary Wrapper. In this case, we have built five different classifiers, each specialized in classifying the example of one class. We formed five training sets, one for each classifier, where there are only two classes: class member or nonmember. Each classifier calculates the probability of each example being a member of each class. Then these probabilities are compared and normalized. Features are selected for each classifier, and we have five feature subsets ($F_1, F_2, F_3, F_4,$ and F_5), each of which has a number of k features. All these feature subsets are built as described above, according to the algorithm described in [15]. So, for each class, we select the feature with the best accuracy classifying whether

or not each example belongs to the class. This feature is added to the subset, and the process starts again. Figure 8 shows the process described above. We have selected the features according to this algorithm, and we have built five feature subsets (one for each class) for one, two, and three cardinalities. Table 2 shows the selected features.

4.4. Global mRMR Filter. In this case, we have selected the subset of features based on the mRMR feature selection algorithm described in [18]. This algorithm uses the mutual information criteria of max-relevance and min-redundancy described in (10). This algorithm selects a subset of k features:

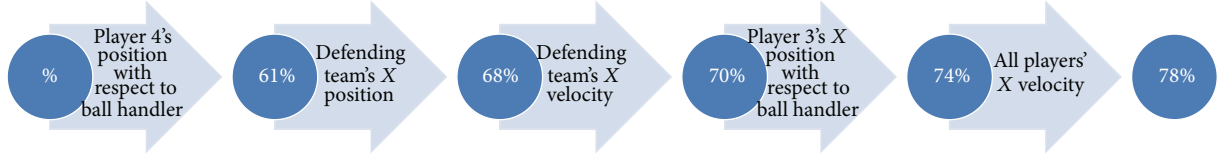$$I(C_i, f_i) - \frac{1}{|F|} \sum_{f_s \in S} I(f_s, f_i). \tag{10}$$
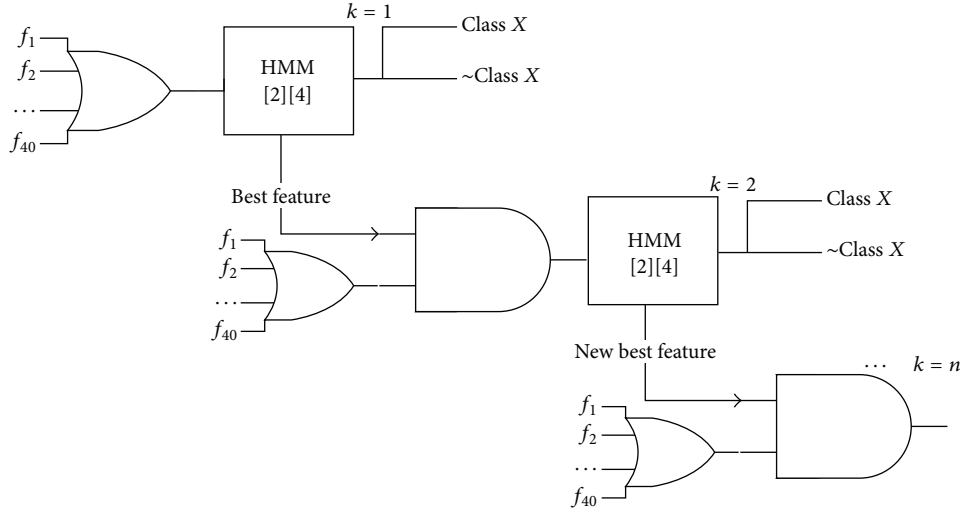
FIGURE 7: Selected features.

FIGURE 8: Binary wrapper algorithm.

TABLE 2: Features selected (BWC).

k	Class	New feature selected	Accuracy
1	1	Player3's X position	70%
	2	Player3's Y position	
	3	Player1's X position	
	4	Player2's Y velocity	
	5	X coordinate of the center of the defending team	
2	1	Player4's X position	91%
	2	Player3's X position	
	3	Player4's X velocity	
	4	Player2's X velocity	
	5	Player4's Y position with respect to the ball handler	
3	1	Y velocity of the attacking team	96%
	2	First component of the Laplacian invariant of the graph	
	3	Player4's X position with respect to the ball handler	
	4	X velocity of the attacking team	
	5	Player2's X velocity	

Figure 9 shows the process described above. We have selected the features according to this algorithm, and we have built a subset for one-to-seven cardinality. Table 3 shows the selected features.

4.5. Binary mRMR Filter. In this feature selection algorithm, we have built five classifiers as described in Section 4.3, and we have calculated five subsets of features (one for each classifier) using the mutual information mRMR criterion described above. So, we have five subsets of features, composed of a number k of features, each specialized in each class.

Figure 10 shows the process. We have selected the features according to this algorithm, and we have built five subsets (one for each class) for one-to-seven cardinality. Table 4 shows the selected features.

5. Experimentation

We have used MATLAB framework for the experiments. The entire algorithm for the behavior recognition issue was expressed in this language. First of all, the dataset was loaded into the system, with all the extracted features and
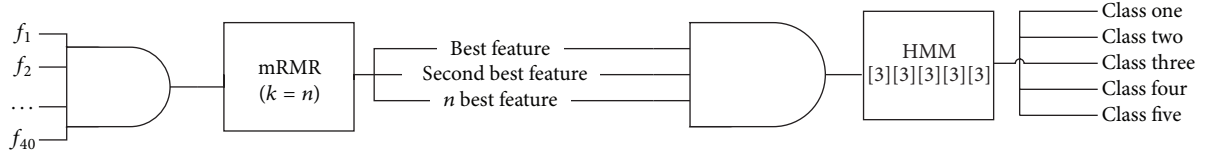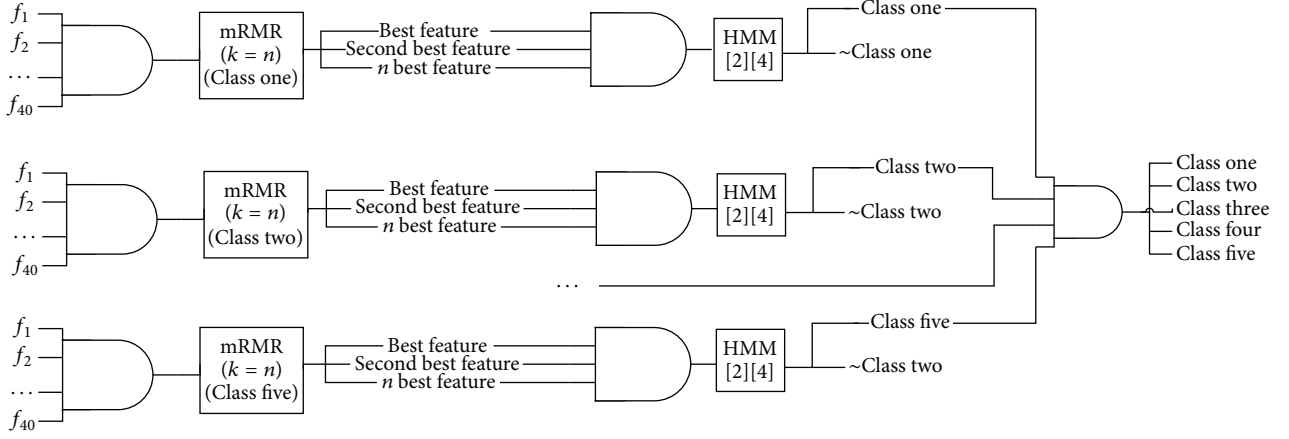
FIGURE 9: Global rMRM filter algorithm.

FIGURE 10: Binary rMRM filter algorithm.

TABLE 3: Features selected (GFC).

k	New feature selected	Accuracy
1	Player1's X position	39%
2	First component of the Laplacian invariant of the graph	61%
3	Third component of the Laplacian invariant of the graph	61%
4	Second component of the Laplacian invariant of the graph	61%
5	Ball handler	61%
6	Y velocity of the defending team	53%
7	Player4's Y position	43%
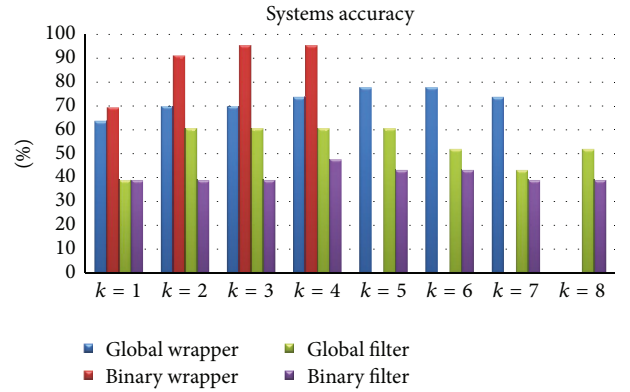
FIGURE 11: Systems accuracy.

data. The INEF12Basketball dataset (which is available at http://www.giaa.inf.uc3m.es/) provides 27 examples of a set of eight different group behaviors. For these experiments, we selected five of these behaviors performed by 23 of the examples. The other behaviors were omitted because there are too few examples. Then we entered all the information, representing all features at all times, into a MATLAB matrices. We have one matrix M_i ($1 \leq i \leq 23$) per group action and one dimension vector containing the behavior label. Each M_i matrix has $F \times T_i$ dimensions, where F is the number of selected features and T_i is the duration (in frames) of behavior i. These matrices are the input for the described algorithms, using leave-one-out cross-validation [22]. Note that not all the frames have been loaded, because only a few seconds per sequence are meaningful for the move. So, only one several-second subsequence selected by the expert has

been loaded per sequence. For the experiments, we used a well-known technique called hidden Markov model. We used a HMM with three Gaussians per class for all experiments with global wrapper and global mRMR filter, and we used two Gaussians for class members and four Gaussians for class nonmembers for binary wrapper and binary mRMR filter. This paper sets out to validate a new representation of group behaviors. On this ground, we have not bothered too much about technique selection. It is assumed that the reported research, both feature selection and representation, is validated if the technique yields good results. Figure 11 shows system accuracy with one, two, three, four, five, six, seven, and eight features. Clearly, binary wrapper yields the best results with a classification accuracy of 96% for only three features per class. Note that the computational cost of

TABLE 4: Features selected (BFC).

k	Class	New feature selected	Accuracy
	1	Player1's Y position with respect to the ball handler	
	2	Player3's X position with respect to the ball handler	
1	3	Player1's X position	39%
	4	Player1's X position	
	5	Player4's Y position with respect to the ball handler	
	1	First component of the Laplacian invariant of the graph	
	2	First component of the Laplacian invariant of the graph	
2	3	First component of the Laplacian invariant of the graph	39%
	4	First component of the Laplacian invariant of the graph	
	5	First component of the Laplacian invariant of the graph	
	1	Third component of the Laplacian invariant of the graph	
	2	Third component of the Laplacian invariant of the graph	
3	3	Third component of the Laplacian invariant of the graph	39%
	4	Third component of the Laplacian invariant of the graph	
	5	Third component of the Laplacian invariant of the graph	
	1	Second component of the Laplacian invariant of the graph	
	2	Second component of the Laplacian invariant of the graph	
4	3	Second component of the Laplacian invariant of the graph	48%
	4	Second component of the Laplacian invariant of the graph	
	5	Second component of the Laplacian invariant of the graph	
	1	Y velocity of the defending team	
	2	All players's Y velocity	
5	3	Ball handler	43%
	4	Ball handler	
	5	Y velocity of the defending team	
	1	X velocity of the attacking team	
	2	Ball handler	
6	3	Y velocity of the defending team	43%
	4	Y velocity of the attacking team	
	5	Ball handler	
	1	Ball handler	
	2	X velocity of the defending team	
7	3	X velocity of the attacking team	39%
	4	Player2's Y velocity	
	5	X velocity of the attacking team	

this algorithm type is huge. On the other hand, mRMR filter-based algorithms managed to provide 61% accuracy only with two features at a much lower computational cost.

Blue bars represent system accuracy with the best combination of the specific number of features, and red bars represent the average accuracy for all possible combinations of the specific number of features. As of the selection of five features, the addition of another feature clearly does not improve system accuracy, which stops the features selection algorithm.

6. Conclusions

In this paper we have developed and tested four systems that use some low-level features (like positioning) to recognize high-level features, namely, group behavior. This approach is applied to the 2-on-2 basketball domain, where there are few elements in the scene and, therefore, few features to analyze. Nevertheless, the approach could be applied to more complex systems, and it could be effective in most potential domains. Note, for instance, that most of the group actions in the basketball domain are performed by only a few players (not all the players must take part in all moves). Static context information is very important for defining the scene setting. For example, in our basketball domain, we know that there are two groups with two members. All this context information depends on the specific domain and would sometimes need to be described by an expert. Also feature selection could be useful in very different domains. For example, it could be very useful for improving the

accuracy of the group behavior recognition system if some features do not contribute any information, have too much noise, and so forth. Additionally, a simpler system (with fewer features) is likely to be faster and more resistant because the hidden Markov model would need fewer Gaussians. Our experiments have tested four feature selection algorithms using the same classification algorithm. The results of the experiments demonstrate that the binary wrapper algorithm has the best scores. This algorithm type has a prohibitive computational cost. The mRMR filter is an option in this case, and we have shown that the global version of this algorithm is a better alternative.

7. Future Work

All research approaches aim at generalization. There are two possible ways of doing this in group behavior recognition: domain and number of elements. This means that all approaches should work in very different domains and with different numbers of elements in the scene. For this purpose, we think that the INEF12Basketball Dataset proposed in [5] should be extended to more domains with varied number of elements. In this way, all approaches could be tested in different domains, and we could examine system scalability. This is very important in the group behavior recognition research field, because some authors believe that this renders some algorithms inapplicable for this issue [13]. Another important aspect that could be improved in this approach is robustness. This could be done by making the system more resistant to tracker loss. For this purpose, we need a system that could manage a variant number of elements in each group. This is potentially a very challenging and useful improvement. In this paper, we have dealt with the problem of classifying a segmented sequence based on group behavior. In our opinion, this is only the first step of a more ambitious issue. In most real-world cases, we do not have segmented sequences, so we need a system that would be able to automatically segment (and classify) group behavior. Obviously, the system must be trained with segmented examples but would then be able to automatically segment a sequence with more than one group behavior.

Acknowledgments

References

[1] J. K. Aggarwal and Q. Cai, "Human motion analysis: a review," in *Proceedings of the IEEE Nonrigid and Articulated Motion Workshop*, pp. 90–102, June 1997.

[2] A. F. Bobick and A. D. Wilson, "A state-based approach to the representation and recognition of gesture," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 12, pp. 1325–1337, 1997.

[3] A. Pozo, J. Garca, and M. Patricio, "INEF12basketball dataset and the group behavior recognition issue," in *Distributed Computing and Artificial Intelligence*, vol. 151 of *Advances in Intelligent and Soft Computing*, pp. 151–160, Springer, New York, NY, USA, 2012.

[4] L. R. Rabiner, "Tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[5] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Computer Vision and Image Understanding*, vol. 104, no. 2-3, pp. 90–126, 2006.

[6] K. Schindler and L. van Gool, "Action snippets: how many frames does human action recognition require?" in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, Anchorage, Alaska, USA, June 2008.

[7] R. Li and R. Chellappa, "Recognizing offensive strategies from football videos," in *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP '10)*, pp. 4585–4588, IEEE, September 2010.

[8] M. Perše, M. Kristan, S. Kovačič, G. Vučković, and J. Perš, "A trajectory-based analysis of coordinated team activity in a basketball game," *Computer Vision and Image Understanding*, vol. 113, no. 5, pp. 612–621, 2009.

[9] A. Pozo, J. Gracía, M. A. Patricio, and J. M. Molina, "A structured representation to the group behavior recognition issue," *Advances in Intelligent and Soft Computing*, vol. 94, pp. 47–57, 2011.

[10] S. Khan and M. Shah, "Detecting group activities using rigidity of formation," in *Proceedings of the 13th Annual ACM International Conference on Multimedia (MULTIMEDIA '05)*, pp. 403–406, ACM, 2005.

[11] F. Cupillard, A. Avanzi, F. Bremond, and M. Thonnat, "Video understanding for metro surveillance," in *Proceeding of the IEEE International Conference on Networking, Sensing and Control*, vol. 1, pp. 186–191, March 2004.

[12] F. Cupillard, F. Brémond, and M. Thonnat, "Behaviour recognition for individuals, groups of people and crowd," in *Proceedings of the IEE Symposium on Intelligence Distributed Surveillance Systems*, vol. 7, pp. 1–5, 2003.

[13] R. Li, R. Chellappa, and S. K. Zhou, "Learning multi-modal densities on discriminative temporal interaction manifold for group activity recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR '09)*, pp. 2450–2457, IEEE, June 2009.

[14] J. Pers, "Cvbase 06 dataset: a dataset for development and testing of computer vision based methods in sport environments," 2006.

[15] M. Patricio and D. Maravall, "Segmentation of text and graphics/images using the gray-level histogram fourier transform," in *Advances in Pattern Recognition*, vol. 1876 of *Lecture Notes in Computer Science*, pp. 757–766, Springer, New York, NY, USA, 2000.

[16] S. Maldonado and R. Weber, "A wrapper method for feature selection using support vector machines," *Information Sciences*, vol. 179, no. 13, pp. 2208–2217, 2009.

[17] B. Fish, A. Khan, N. Hajj Chehade, C. Chien, and G. Pottie, "Feature selection based on mutual information for human activity recognition," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '12)*, pp. 1729–1732, Kyoto, Japan, March 2012.

[18] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern*

Analysis and Machine Intelligence, vol. 27, no. 8, pp. 1226–1238, 2005.

[19] H. Liu, J. Sun, L. Liu, and H. Zhang, "Feature selection with dynamic mutual information," *Pattern Recognition*, vol. 42, no. 7, pp. 1330–1339, 2009.

[20] W. T. Chen, K. F. Ji, X. W. Xing, H. X. Zou, and H. Sun, "Ship recognition in high resolution sar imagery based on feature selection," in *Proceedings of the International Conference on Computer Vision in Remote Sensing (CVRS '12)*, pp. 301–305, Xiamen, China, December 2012.

[21] K. Ausfeld, Z. Ninkov, P. P. Lee, J. D. Newman, and G. Gosian, "Polynomial fitting adaptive kalman filter tracking and choice of correlation coefficient," in *26th Acquisition, Tracking, Pointing, and Laser Systems Technologies*, vol. 8395 of *Proceedings of SPIE*, Baltimore, Md, USA, April 2012.

[22] R. Picard and R. Cook, "Cross-validation of regression models," *Journal of the American Statistical Association*, vol. 79, no. 387, pp. 575–583, 1984.

The Scientific World Journal

Hindawi

- ▸ Impact Factor **1.730**
- ▸ **28 Days** Fast Track Peer Review
- ▸ All Subject Areas of Science
- ▸ Submit at http://www.tswj.com