# Semantic Categorization of Video: An Evolutionary Learning Based Approach


| R.S. Jadon | Santanu Chaudhury | K.K. Biswas | Asma Shakil |
|---|---|---|---|
| CSE Dept. | EE Dept. | CSE Dept. | EE Dept. |
| IIT Delhi | IIT Delhi | IIT Delhi | IIT Delhi |
| New Delhi-16 | New Delhi-16 | New Delhi-16 | New Delhi-16 |
| rsjadon@cse.iitd.ernet.in | santanuc@ee.iitd.ernet.in | kkb@cse.iitd.ernet.in | asmashakil@hotmail.com |



### Abstract

In this paper, we have proposed a fuzzy rule based system for classification of video into semantic categories. The classification scheme uses an evolutionary learning methodology to evolve a fuzzy system for use in the classification process. This evolved fuzzy classifier has the inherent capability to tackle variations and ambiguities invariably present in the video data. A novel fuzzy theoretic scheme has been suggested for extraction of key frames from a given video after shot segmentation. Frame based temporal features and spatial features obtained from key frames have been used in the classification system. We have developed an experimental system for categorization of sports video. The experimental system has yielded reasonably correct recognition results for a large number of samples.


## 1 Introduction

Information and entertainment appliances delivering video according to user's preferences can become reality if powerful methods are developed for content characterization of video. In this paper, we have proposed a fuzzy rule based system for classification of video into semantic categories. The classification scheme uses an evolutionary learning methodology to evolve a fuzzy system for use in the classification process. This evolved fuzzy classifier has the inherent capability to tackle variations and ambiguities invariably present in the video data.

Different approaches have been suggested in the past for semantic categorization of video sequences. Zhang et al.[20] have used models of specific types of programs such as TV news. Model based or knowledge based approaches have been also proposed for analysis of sports video like soccer [4] and tennis[13]. Use of a rule based system has been proposed in [22] for categorization of basketball video into different classes using visual and motion based characteristic features. They have used an inductive decision-tree learning method to arrive at a set of if-then rules. Caliani et al.[1] suggested use of higher level semantic features for categorization of commercial video into semantic categories. Szummer and Picard [15] suggested use of features computed on image sub-blocks for classification of images as indoor and outdoor scenes. A three-level video-event detection methodology and its application to hunt-event detection have been described in [6]. The first level extracts color, texture and motion features and moving object blobs. The mid-level uses neural network based object classification module. A domain specific inference process is employed at the third level for event detection. As expected, all these semantic categorization schemes exploit domain knowledge. Some of the systems also employ different paradigms of learning [22, 6] for designing the classifiers. However, these schemes have problems when training data has inherent variability and is noisy[10]. In order to overcome this problem, in this paper, we have proposed use of an evolutionary learning based fuzzy rule based classifier. The fuzzy rules can take care of the inherent impreciseness of the feature set for the classification task. Further, the evolutionary learning scheme has the ability to deal with noisy training data for extraction of apparently inconsistent domain dependent fuzzy rules which are applicable under different conditions.

We have proposed a multi-tiered classification architecture. We have used learning and fuzzy rules for implementation of each layer (see Fig.1) . At the first level we have used fuzzy rules for video segmentation, as reported in our earlier work[7]. This scheme provides a mechanism for characterizing scene transitions through soft-decisions. Subjectivity involved in the video partitioning process can be adequately captured by this framework. This scheme can clearly distinguish between abrupt and gradual transitions. At the next level, we have used another fuzzy rule based system for extraction of key frames abstracting each shot. This is a novel contribution of this paper. At the subsequent level, the video is represented in terms of fuzzy features extracted from frame sequences and/or key frames. We have proposed use of feature-spaces which can capture semantics of video sequences in terms of dynamics of scenes and presence or absence of a fuzzy object. In fact, fuzzy theoretic specification of the image based features and spatial relations between image components have the ability to accommodate variabilities inherent in the individual categories of the video sequences. A set of fuzzy rules have been learnt using training examples for detection of these features. At the top level, we have a fuzzy rule based categorization system. This system uses fuzzy features extracted in the previous layer for semantic categorization. The same learning paradigm has been adopted for evolving this classification system. This multi-layered classification architecture where each module can be trained individually is the fundamental contribution of this work. We have considered the domain of sports for implementation of the classification scheme. Through learning we have evolved rule modules for extracting features relevant for the domain of tennis, football, cricket, tennis and sprinting. Further, we have learnt rules for classification of a sports video into one of these classes. Experimental results have established effectiveness of this approach.

The paper is organized as follows: we briefly discuss evolutionary learning paradigm in section 2. In section 3, the segmentation and key-frame extraction scheme is presented. In section 4 we discuss the fuzzy feature extraction using evolutionary fuzzy system. The overall fuzzy rule based video classification scheme for sports video is presented in section 5. The experimental results are given in section 6. Finally we conclude in section 7.

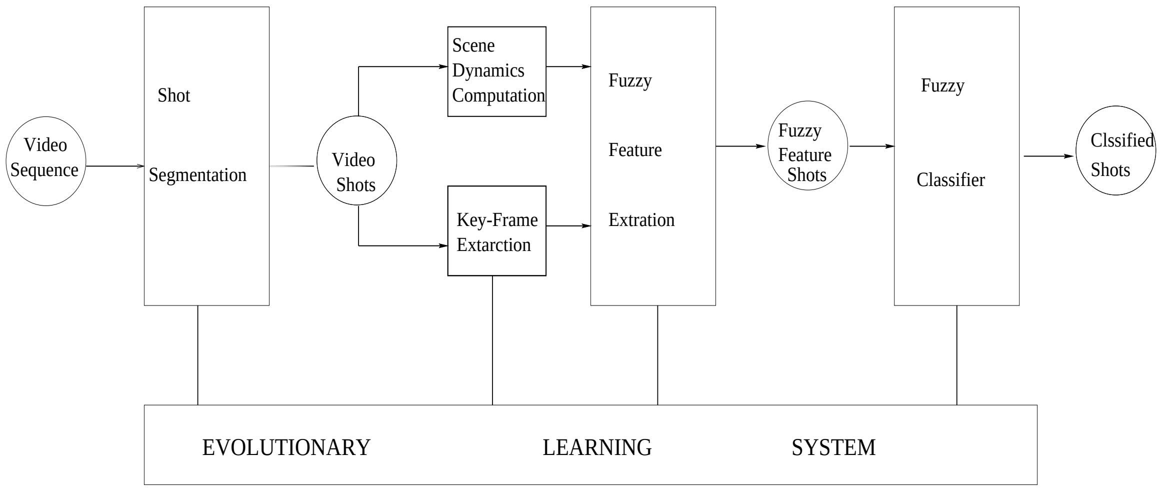



Figure 1: An Overview of Video Characterization System

## 2 Evolutionary Learning of Fuzzy Rule Based System

The fuzzy rule based systems used at different levels of our classification architecture has been learnt using an evolutionary learning scheme. The learning scheme is proposed in [12]. In this section, for the sake of completeness, we are briefly reviewing the approach suggested in [12].

The evolutionary learning scheme involves following issues:

- Designing of an encoding scheme for representation of the fuzzy rule based system in terms of chromosomes
- Generation of initial population
- Calculation of fitness of each chromosome in the population and selection of chromosomes for reproduction
- Reproduction of new chromosomes using crossover and mutation operation.

The steps of fitness value computation, selection and reproduction is iterated until convergence condition is met.

To completely represent a fuzzy system, each chromosome must encode all the needed information about the rule set and the membership functions. As an example Table 1 shows a fuzzy system that can be used for classification of sports video. This fuzzy system has three

| | | | |
|---|---|---|---|
| input | green_score | | |
| input | pitch_score | | |
| input | player_score | | |
| output | sport-type | | |
| | cricket | football | golf |

Table 1: An Example of a Fuzzy Expert System

input variables - green-score, pitch-score and player-score and one output variable - sport-type. Each input variable has three fuzzy sets representing the linguistic descriptions: *low, medium* and *high*. We can use the integers 1,2 and 3 to represent each of these three terms. The output variable can have three possible values shown in the last row of the Table 1 and can be represented by three integers. A fuzzy rule can now be completely represented by four integers. For example , the rule *if green-score is high AND pitch-score is low AND player-score is low THEN sport-type is Golf* can be encoded as 3111. If the rule set includes 20 rule, then an integer string of length 80 can represent the rule set completely. The membership functions and class boundaries can also be represented as integers [12]. Typically length of the string is determined on the basis of a guess about the maximum possible number of rules. The complete fuzzy system thus can be encoded as a string of integers. This string of integers represent a chromosome for the genetic algorithm-based learning system.

The next important consideration following the representation is the choice of the fitness function. The genotype representation encode the problem into a string while the fitness function measures the performance of the system. To find a good fitness measurement is quite important for evolving practical systems using GA's. The fitness function we have chosen consists of assigning to each individual (i.e. each FRBS) a reward in proportion to the number of correctly classified patterns and the degree of correctness. For crossover operation two randomly generated partitions of the string are exchanged. Mutation involves alteration of integer values within the permissible range. In the present implementation the probabilities of crossover and mutation are not held constant for the entire run of GA, instead they are varied during the run, using a set of eight fuzzy rules as proposed in [12].

| Quality Measure Measure | Hand-Crafted FRBS | Evolved FRBS |
|---|---|---|
| Accuracy | 86.47 | 88.89 |
| Precision | 92.73 | 94.30 |
| Recall | 98.88 | 100 |
| Error R. | 6.26 | 4.43 |

Table 2: Performance Evaluation Results for Fuzzy Segmentation

# 3 Shot Segmentation and Key Frame Extraction

Since shots are a fundamental unit of video, therefore as a first step, the video sequence to be characterized is segmented into shots. Using evolutionary learning fuzzy rule based system is developed for this purpose, which is an improvement over our earlier work reported in[8]. Here we need to provide only the training patterns in the learning phase, which evolves the FRBS for video sequence segmentation. A comparative study of hand crafted FRBS and evolved FRBS with regard to performance measures as described in [11] for segmentation of the video sequence into shots is given in table2. The fuzzy labels: *negligible*, *small*, *large* and *very-large*, computed on the basis of the frame-to-frame RGB histogram intersection values, are associated with each frame-transition during the segmentation phase. These labels are further used for key-frame extraction.

## 3.1 Key Frame Extraction

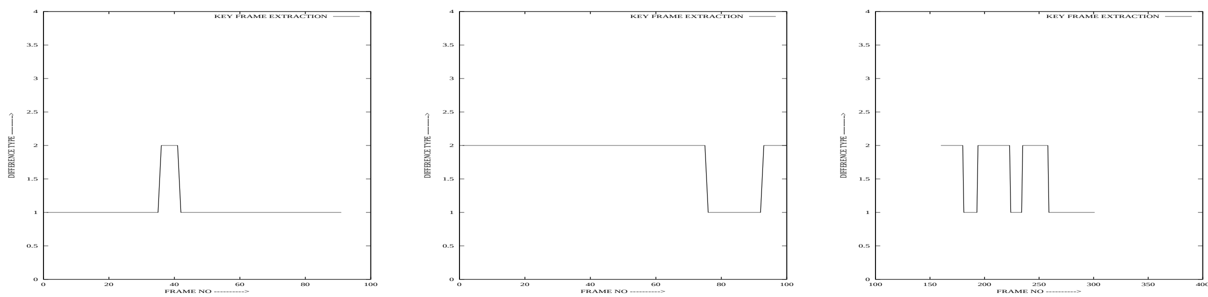
Representation of video content is a difficult problem. Many schemes are suggested in the literature for this purpose. Tonomura et al. [17] use the first frame of each shot as a key-frame. Ueda et al. [18] represent shots with two key-frames - the first and last frames of each shot. Ferman et al.[2] use clustering on the frames within each shot. The frame closest to the center of the largest cluster is selected as the key frame for that shot. Taniguchi et al.[16] generate a composite image to represent shots with camera motion. Zhang et al.[21] and Gunsel et al.[5] segment the video into shots and select the first clean frame of each shot as a key frame. Yeung et al.[19] select one key frame frame for each shot. These key frames are then clustered based on visual similarity and temporal distance. Sun et al. [14] divide a video into intervals with the largest dissimilarity between the first and last frame. The key-frame selection technique given by Andreas et al.[3] makes use of clustering and temporal constraints. An innovative technique using object-based video abstraction is proposed by Changick et al[9]. The object-based key-frame selection schemes are highly task specific and can be used only in restricted domains. The clustering based technique as suggested by [3] and [2] seems to be more appealing for our application, but the major drawback of most of these techniques is deciding thresholds for the size of clusters and temporal constraints. To overcome these problems we have developed a temporal-clustering-based fuzzy key-frame extraction technique. Logically frames are clustered depending on the type of frame-to-frame transition as indicated by the segmentation module. The number of consecutive frames having the same transition label is fuzzified to obtain the fuzzy input variable **Size-of-cluster**. The other input variable is the **Type-of-cluster** and its membership value is obtained as the maximum of the membership value of individual frame-to-frame transitions. The fuzzy rule based system is evolved using evolutionary learning paradigm as given in the previous section. Some of the typical rules evolved are given below:

- If CLUSTER-TYPE is small and CLUSTER-SIZE is large then it is SINGLE-KEY-CLUSTER
- If CLUSTER-TYPE is small and CLUSTER-SIZE is small then it is NON-KEY-CLUSTER
- If CLUSTER-TYPE is small and CLUSTER-SIZE is very-large then it is SINGLE-KEY-CLUSTER
- If CLUSTER-TYPE is large and CLUSTER-SIZE is large then it is SINGLE-KEY-CLUSTER
- If CLUSTER-TYPE is large and CLUSTER-SIZE is small then it is NON-KEY-CLUSTER

- If CLUSTER-TYPE is large and CLUSTER-SIZE is very-large then it is MULTIPLE-KEY-CLUSTER

We classify the clusters into three categories: non-key-cluster, single-key-cluster and multiple-key-cluster. Non-key clusters are small size clusters not carrying any significant information and thus can be ignored. Single-key-clusters are large-sized clusters from which we extract the centroid as the key-frame. Multiple-key-clusters are very-large-size clusters where frame-to-frame transitions are in general *large*. This can be due to panning, zooming or high-object-motion. These clusters will be represented by more than one representative frames. This heuristic intuitively removes redundancies retaining only relevant frames. Multiple-key-clusters are further divided into two equal sized clusters and the rules are applied again. This processes is iterated until a multiple-key-cluster is divided into many single-key clusters. A key-frame is then extracted from each cluster. A typical plot of defuzzified frame-to-frame difference-values wrt. Frames-no. is shown in Fig.2. In these figure we have represented the *small-difference* with *1* and large difference with *2*. The x-axis corresponds to the frame numbers and y-axis to difference values. In Fig.2(a) there are two single-frame-clusters, separated by one non-key-cluster. In Fig.2(b) there is one multiple-key-frame-cluster, one single-key-frame-cluster and one non-key-cluster. In Fig.2(c) there are three single-frame-clusters (two are of large-transition-type and one of small-transition-type).

A typical set of key-frames from a panning cricket video shot is shown in Fig. 3. In this sequence three representative frames have been selected corresponding to small-transition-type clusters of large size. The sub-sampled sequence is shown in Fig.4. For shots with high camera motion or object motion it becomes difficult, even for the human observer, to select the key frames correctly. We have performed a subjective test (with 10 observers) to test the correctness of our algorithm. In 95% cases, users were satisfied with the set of key frames obtained by our algorithm. In the remaining cases observers reported redundancies in the key-frame set. However, observers did not report non-inclusion of desirable frames. The problems were primarily in those shots which contained a high degree of combination of camera effects and/or object motion. One such sub-sampled video is shown in Fig.5. The set of key frames selected for this video is shown in Fig.6. In this shot a player is running




(a) (b) (c)

Figure 2: Frame difference Vs. Frame-no (a)Normal (b)Single (c)Multiple

(a) (b) (c)

Figure 3: Key Frame Set for4:(a)Frame-15 (b)Frame-25 (c)Frame-48

behind the ball and camera is panning to capture the action of the player. While the player is running parallel to the boundary, the advertisement and posters in the background are changing rapidly. Consequently, many large regions of large-transition-type of clusters are generated, giving redundant frames in the the key frame set (see Fig.6).

# 4 Fuzzy Feature Extraction

Fuzzy feature extraction is an important activity for semantic content characterization. We have extracted spatial and temporal features for each shot. Spatial features are extracted from key-frames, while domain-knowledge-based scene dynamic features are extracted from the frame sequences of the complete shot. In the subsequent subsections we will discuss the scheme for extracting these features.


Figure 4: A Sub Sampled Cricket Shot(Panning)

Figure 5: Running Player against changing background

Figure 6: Key-Frames for Running Player Shot

## 4.1 Key-Frame Based Features

The high-level fuzzy semantic features are extracted from key-frames by using low-level image based features. We have used HSI color space model for extracting pixel-level features. The HSI-model best fits into the criteria of color perception because as long as the colors in the image are reasonably well saturated hue will tend to remain relatively constant in the presence of shadows and other lighting variations. In such cases an image based on hue alone may work better than traditional gray scale analysis. We segment the key-frame in terms of HSI sub-space. We than compute various low-level features for regions such as centroid, area, perimeter, moment invariants, color etc. A fuzzy rule based system is then used for classification of the regions into semantically meaningful entities. The evolutionary learning scheme is used for learning these rules.

Semantically meaningful entities are identified on the basis of the domain knowledge. Fuzzy rules are learnt using manually identified examples of these entities. In the following we have described in detail feature extraction schemes developed for our chosen domain of sports comprising of five disciplines: cricket, football, golf, tennis and sprinting. Some of these features are general, some others are game specific. However, for an unknown frame, each region (greater than a minimum size - measured in terms of the number of pixels) is analyzed by the same set of fuzzy rules for possible classification.

**GREEN SCORE:** The variable green score is a measure of green content in the given input image. The hue value for green is about 120 degrees from the red axis in the HSI model. This feature is used to calculate the green score. First the image is transformed into HSI coordinates then we count the pixels associated with green color. This count is normalized wrt the dimensions of the image. We call this as *normalized-green-pixel-count*. It is taken as input variable for the purpose of fuzzification. The output variable *green-score* is determined from this input variable. We associate three fuzzy sets to the input variable *normalized-green-pixel-count*: small, large and very-large. Similarly three sets are associated with the output variable *green-score*: low, medium and high. We then train the system using evolutionary learning as discussed in the previous section to evolve the FRBS. The evolved membership functions and class boundaries for the input variable is shown in the Fig.8. Some of the evolved fuzzy-rules are as shown below:

- If *normalized-green-pixel-count* is **small** then *green-score* is **low**
- If *normalized-green-pixel-count* is **large** then *green-score* is **medium**
- If *normalized-green-pixel-count* is **very-large** then *green-score* is **high**

**PITCH SCORE:** The pitch score for a region in the image is calculated using three of the attributes i.e. pitch-shape, pitch-color and surroundedness with green color.

**Pitch Shape:** It is characterized as a rectangular region. We have computed seven moment invariants for sample pitch regions. A mean moment vector is obtained from these examples. For a given region, difference of its moment vector with the mean vector is fuzzified. This input variable is called *error-value*. The corresponding output variable for the fuzzy system is characterized as the *pitch-shape-score*. The training set thus consists of *error-value* and the *pitch-shape-score)* characterized by that value. The fuzzy-input variable *error-value* is characterized by three input sets: small, large and very large. Similarly fuzzy-output variable pitch-shape-score is characterized by fuzzy-sets low, medium and high.

**Pitch Color:** Pitch color characterized by the count of brownish-yellow pixels in the region. Its computation is similar to the green-score computation for the image. Here we obtain *pitch-color-score* as the fuzzy-output.

**Green-Surroundedness-score:** For surroundedness calculation we have taken a distance of 10 pixels in four directions from the centroid of the region. The Eucledian distance between a green pixel and the pixel which is 10 pixels apart from the centroid of the region is computed in all the four directions. We compute the mean of this distance and use it as input variable for our evolutionary fuzzy system. The fuzzy sets associated with this distance are: small, large and very-large. The fuzzy output variable is *green-surroundedness-score* having fuzzy sets low, medium and high. The fuzzy system is trained with appropriate data and the resulting evolved fuzzy-rule-based system is used to calculate the *green-surroundedness-score* for the region.

Finally the fuzzy attribute **PITCH-SCORE** is determined using these attributes. At this stage we evolve one more fuzzy rule based system using pitch-shape-score, pitch-color-score and green-surroundedness-score as input fuzzy-variables and pitch-score as the output variable. Some of the evolved fuzzy rules for pitch-score computation are as shown below:

- If *pitch-color-score* is **high** and *pitch-shape-score* is **high** and *green-surroundedness-score* is **high** then *pitch-score* is **high**
- If *pitch-color-score* is **high** and *pitch-shape-score* is **low** and *green-surroundedness-score* is **high** then *pitch-score* is **medium**
- If *pitch-color-score* is **high** and *pitch-shape-score* is **high** and *green-surroundedness-score* is **low** then *pitch-score* is **medium**
- If *pitch-color-score* is **low** and *pitch-shape-score* is **high** and *green-surroundedness-score* is **low** then *pitch-score*e is **low**
- If *pitch-color-score* is **medium** and *pitch-shape-score* is **high** and *green-surroundedness-score* is **high** then *pitch-score*e is **high**
- If *pitch-color-score* is **medium** and *pitch-shape-score* is **medium** and *green-surroundedness-score* is **high**h then *pitch-score*e is **medium**
- If *pitch-color-score* is **medium** and *pitch-shape-score* is **medium** and *green-surroundedness-score* is **medium** then *pitch-score*e is **medium**

The evolved fuzzy membership functions and class-boundaries are shown in Fig.7.

**PLAYER-SCORE:** The fuzzy rules for *player-score* are obtained using an approach similar to pitch score. We use shape of the player region because that corresponds to an elliptical region for a distant view. Also we exploit the fact that a player region must be surrounded by a green region.

**COURT SCORE:** The court is used as the key feature for tennis video. For this, first the court lines are detected using color information. Next Hough Transform is used to determine straight line configurations in the segmented court region. Once the lines have been detected, fuzzy rules are used to decide whether the detected lines characterize a tennis-court shape or not. The low-level fuzzy predicates used in computing the *court-score* are pseudo-parallel-vertical-pair, pseudo-parallel-horizontal-pair, inter-vertical-line and separation between these pairs. These attributes are computed from the count of pixels for straight-line configurations detected using Hough Transform. These low-level fuzzy attributes characterize intermediate-level fuzzy attributes like *vertical-court-boundary and horizontal-court-boundary.* Finally the evolutionary fuzzy system is designed to extract the *court-score.* Some of evolved fuzzy rules for characterizing the *court-score* are shown below:

- If *pseudo-parallel-vertical-pair* is small AND *separation* is large AND *inter-vertical-line* is large THEN *Vertical-Court-Boundary-Score* is high
- If *pseudo-parallel-horizontal-pair* is small AND if *separation* is moderate AND *inter-horizontal-line* is large THEN *Horizontal-Court-Boundary-Score* is high
- If *Vertical-Court-Boundary-score* is small AND *Horizontal-Court-Boundary-score* is large THEN *court-score* is high

## 4.2 Video Characterization Based On Scene Dynamics

Dynamic changes of object positions also provide an important clue for video characterization. In the present work, we have exploited the dynamic information present over different frames of a sports video to characterize the change in the configuration of players across different frames. For this, a polygon is constructed with the players as its vertices. Next the change in the shape of the polygon across different frames is computed in terms of the difference in the moments of the polygon shape. For sequences with fast relative motion between the players

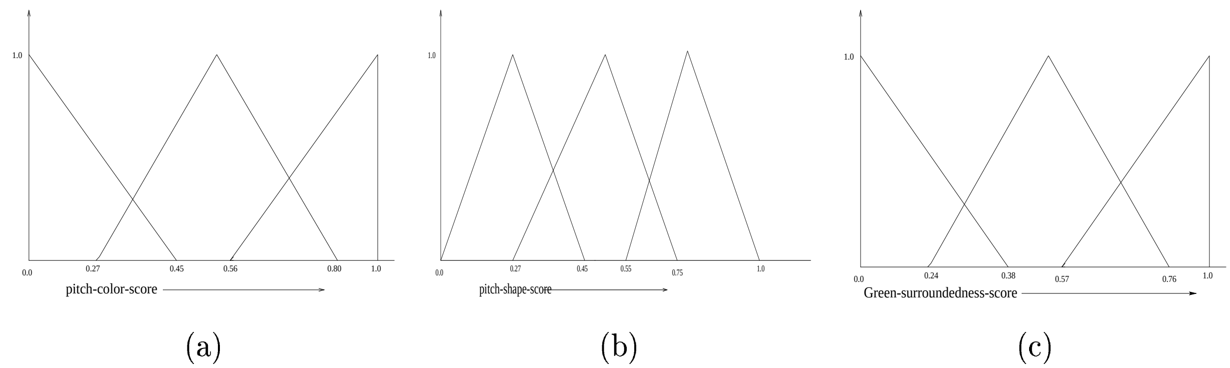



Figure 7: Evolved Membership Functions for(a)pitch-color-score (b)pitch-shape-score(c)green-surroundedness-score

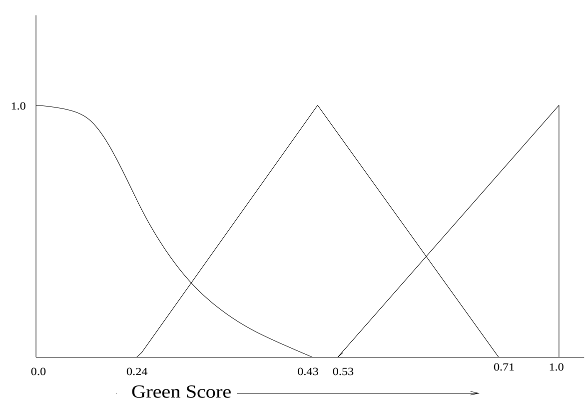



Figure 8: Evolved Membership Functions and Class Boundaries for Green-score

(e.g. football sequences) the shape of the polygon changes rapidly while the change is slow for rather sedate games (like cricket). Motivated by this knowledge we have fuzzified the **spatial configuration change** so that it can be categorized as a **'no-change' or 'a large slow change' or 'a small slow change' or 'a large fast change' or 'a small fast change'**. The following features are chosen for this purpose:

- Normalized range of difference values so that it can be labeled as small, medium, large and very large.

- Normalized coefficient of variation of the difference values which is labeled as being either small, medium, or large.

Some typical rules for video classification based on spatial configuration change are given below:

- *Rule1: If range is SMALL and coefficient of variation is SMALL then the sequence is UNCHANGED*

- *Rule2: If range is MEDIUM and coefficient of variation is SMALL then the sequence is UNCHANGED*

- *Rule3: If range is MEDIUM and coefficient of variance is MEDIUM then sequence is SLOW CHANGE*

- *Rule4: If range is LARGE and coefficient of variation is MEDIUM then the sequence is SLOW CHANGE*

- *Rule5: If range is large and coefficient of variation is LARGE then the sequence is "FAST CHANGE"*

- *Rule6: If range is VERY LARGE and coefficient of variation is LARGE then change is FAST CHANGE*

We further compute the error between the first and last frame of the sequence, we call this cumulative difference. . This is also fuzzified. Finally the following rules are used to determine the overall dynamics of the scene:

- *If SLOW-CHANGE is Large and CUMULATIVE-DIFFERENCE is Small then SMALL-SLOW-CHANGE*
- *If FAST-CHANGE is Large and CUMULATIVE-DIFFERENCE is Small then SMALL-FAST-CHANGE*

# 5 Fuzzy Rule based Video Classification

Using the fuzzy features as discussed in the previous section, we have designed a evolutionary fuzzy rule based system for categorization of sports video. We have considered rules in the form of fuzzy implications involving conjunctive combination of the fuzzy predicates constructed with these fuzzy variables. Initial population for the evolutionary system consisted of strings representing different membership functions for values of fuzzy variables, different class boundaries and different rules formed by random combinations of fuzzy predicates. Starting with these type of initial population final fuzzy system was evolved. Membership functions and class definitions of these fuzzy variables were learned through the evolutionary learning scheme. We have found the system to evolve and generate rules which were intuitively consistent. For example, rules for categorizing cricket video suggested presence of a predominant field with a certain degree of green color, a pitch somewhat rectangular in shape and players in the ground. Some of the evolved fuzzy rules for sports video characterization are as shown below:

- `Rule1: If (Green-score is high) AND (Pitch-score is low) AND (player-score is low) THEN` *GOLF* `sequence.`

- `Rule2: If (Green-score is medium) AND (Pitch-score is low) AND (player-score is low) THEN` *GOLF* `sequence.`

- `Rule3: If (Green-score is high) AND (Pitch-score is low) AND (player-score is large) THEN` *FOOTBALL* `sequence.`

- `Rule4: If (Green-score is medium) AND (Pitch-score is negligible) AND (player-sco is large) THEN` *FOOTBALL* `sequence.`

- `Rule5: If (Green-score is small) AND (Pitch-score is negligible) AND (player-scor is large) THEN` *FOOTBALL* `sequence.`

- `Rule6: If (Green-score is high) AND (Pitch-score is large) AND (player-score is small) THEN` *CRICKET* `sequence.`

- `Rule6: If (Green-score is high) AND (Pitch-score is large) AND (player-score is small) THEN` *CRICKET* `sequence.`

- `Rule7: If (Green-score is high) AND (Pitch-score is large) AND ( player-score is large) THEN` *CRICKET* `sequence.`

- `Rule8: If (Green-score is high) AND (Pitch-score is medium) AND ( player-score is small) THEN` *CRICKET* `sequence.`

- `Rule9: If (Green-score is high) AND (Pitch-score is medium) AND ( player-score is large) THEN` *CRICKET* `sequence.`

- `Rule10: If (Green-score is medium) AND (Pitch-score is large) AND ( player-score is large) THEN` *CRICKET sequence.*

- `Rule11: If (Green-score is high) AND (Court-score is high) AND ( player-score is low) THEN` *TENNIS sequence.*

- `Rule12: If (Green-score is high) AND (pseudo-parallel-horizontal-pair-score is high) AND ( player-score is high) THEN` *SPRINT sequence.*

- `Rule13: If (Green-score is high) AND (pseudo-parallel-vertical-pair-score is high) AND ( player-score is high) THEN` *SPRINT sequence.*

The spatial rule based system gives satisfactory results for golf, tennis and sprint, while its not that good in distinguishing between football and cricket. To further characterize football and cricket sequences we use dynamic feature based rules. We then evolved the integrated FRBS for characterizing sports sequences using spatial and temporal features. Some of the typical rules for this purpose are as listed below:

- If CRICKET is Large and FOOTBALL is Small and SMALL-SLOW-CHANGE is Large Then FINAL-CRICKET

- If CRICKET is Large and FOOTBALL is Large and SMALL-SLOW-CHANGE is Small Then FINAL-FOOTBALL

- If CRICKET is SMALL and FOOTBALL is Large and SMALL-SLOW-CHANGE is Small Then FINAL-FOOTBALL

- If FOOTBALL is Large and CRICKET is Small and SMALL-FAST-CHANGE is Large Then FINAL-FOOTBALL

# 6 Experimental Results

We have carried out experimentations with 100 video sequences obtained from different television channels. In this section we present some of the experimental results.

First, we present experimental results to establish effectiveness of our scene dynamics based features. The graphs characterizing the scene dynamics for example football and cricket sequences are shown in Fig.9. From the graphs it is clear that there is a large variation in

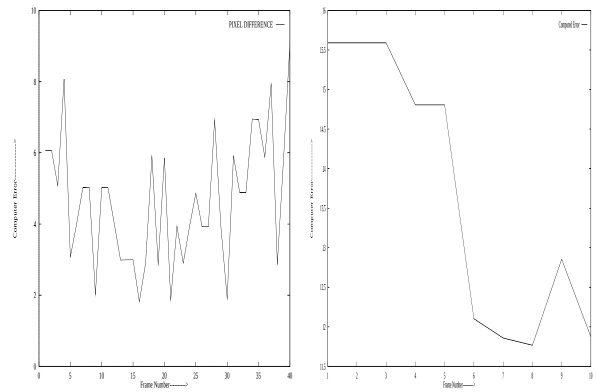


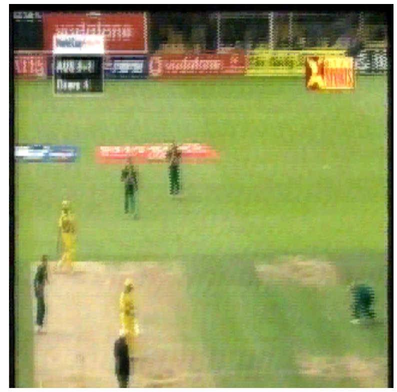

Figure 9: Graph showing error for (a)Football sequence (b)cricket sequence

Table 3: Results Of Fuzzy Characterization of Spatial Configuration Change

| S.No | Movie | Spatial Configuration Change |
|---|---|---|
| 1 | cricket | small slow change with membership 0.237755 |
| 2 | cricket | small slow change with membership 0.335706 |
| 3 | football | large fast change with membership 0.221085 |
| 4 | football | small fast change with membership 0.250414 |
| 5 | football | small fast change with membership 0.149151 |

the moment difference for the football sequence(frames) as compared to that of cricket. In other words we can say that,in general, in a football sequence the spatial configuration change between players is more than in a cricket sequence. Table 3 shows the results of fuzzy analysis of scene dynamics in different video sequences.

Now, we have presented some of the classification results. The fuzzy evaluation scores for two of the cricket key-frames are shown in Fig.10. In Fig.11 the classification score for a football key-frame is shown, along with the key frame. The result for a golf key-frame classification is shown in Fig.12. In Fig.13 the feature extraction and final classification result are shown. Fig.14 shows the results for sprint sequence along with the extracted key-features. In about 80% cases the video was categorized correctly.



Green-Score:0.602033, Pitch-Score: 1.135170
Player-Score: 8.000000

FINAL SCORE (evolved FRBS) :-

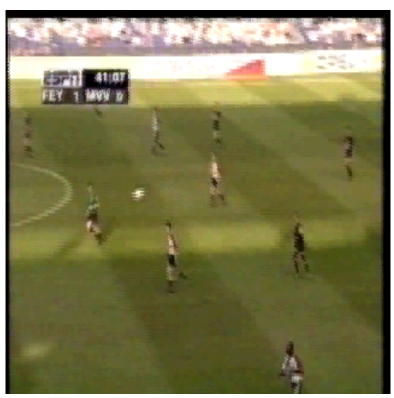
Golf:0.848700 Football:0.999902 Cricket:1.000000

Green-Score: 0.276881, Pitch-Score:0.930405
Player-Score:10.000000

FINAL SCORE (evolved FRBS) :-

Golf:0.346534 Football:0.997885 Cricket:0.999998

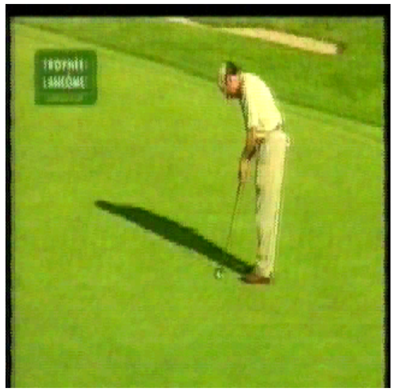
Figure 10: Results of Fuzzy Evaluation for Cricket Frames



Green-Score:0.372682 , Pitch-Score:0.000000, Player-Score:2.000000

(evolved FRBS): Golf: 0.494490 Football: 1.000000 Cricket:0.000030

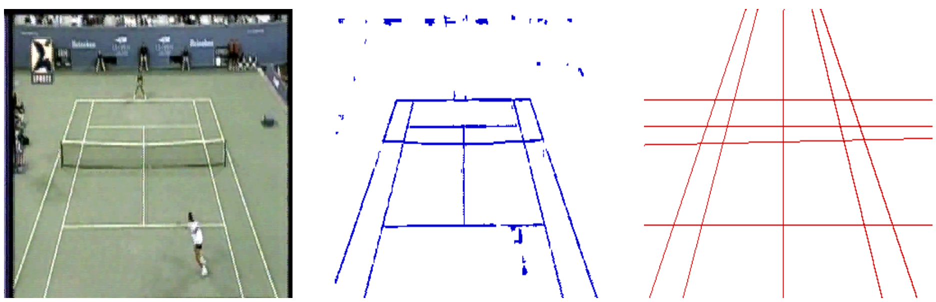
Figure 11: Results of Fuzzy Evaluation for a Football Frame



Green-Score:0.940365, Pitch-Score: 0.926430, Player-Score: 5.000000

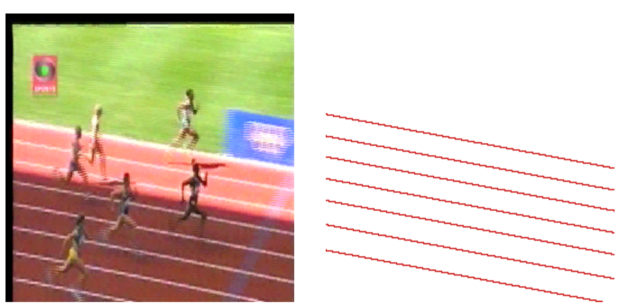
(evolved FRBS): Golf:0.999998 Football:0.842162 Cricket:0.842162

Figure 12: Results of Fuzzy Evaluation for a Golf Frame



(a) (b) (c)

Figure 13: Results of Court Detection In Tennis: (a)Tennis Frame (b)Segmented Court (c)Lines Detected



(a) (b)

Figure 14: Results of Line Detection In Track&Field

# 7 CONCLUSIONS

In this paper we have proposed a generic fuzzy scheme for semantic categorization of video sequences. Starting with the problem of segmenting the video data we have suggested schemes upto content based classification of video sequences. We have primarily focussed on domain-based video categorization problems and chosen the sports video as application domain. The use of the evolutionary learning approach has automated the fuzzy rule based system design. Also, it has provided a mechanism for improving performance of the classification system with the use of additional training samples.

# References


[1] M. Caliani, C. Colombo, and A. Del Bimbo. Commercial Video Retrieval by Induced Semantics. In *Proc. IEEE International Workshop on CAIVD, Bombay, India*, 3 Jan 1998.

[2] A.M. Ferman and Murat Tekalp. Multiscale Content Extraction and Representation for Video Indexing. In *Proceedings of Multimedia Storage and Archiving System II*, volume SPIE 3229, Nov. 1997.

[3] Andreas Girgensohn and John Boreczky. Time Constrained Keyframe Selection Technique. *IEEE Multimedia Systems'99 IEEE Computer Society*, 1:756–761, 1999.

[4] Y. Gong, C.H. Chuan, H.Zhang, and M. Sakauchi. Automatic Parsing of Tv Soccer Programs. *IEEE Transations on Multimedia*, pages 167–172, 1995.

[5] B. Gunsel, Y. Fu, and A.M. Tekalp. Hierarchical Temporal Video Segmentation and Content Characterization. In *Multimedia Storage and Archiving System II, Proc. SPIE 3229, Dallas Tx*, pages 46–56, 1997.

[6] Niels Haering, Richard J. Qian, and M. Ibrahim. A Semantic Event-Detection Approach and Its Application to Detecting Hunts in Wildlife Video. *IEEE Transactions on Circuits and Systems for Video technology*, 10:857–868, September 2000.

[7] R.S. Jadon, Santanu Chaudhury, and K.K. Biswas. Video Segmentation Using Fuzzy Logic. In *Proc. of the Indian Conference on Computer Vision, Graphics and Image Processing(ICVGIP'98), IIT Delhi, Dec'98*, pages 364–370, 21-23 Dec 1998.

[8] R.S. Jadon, Santanu Chaudhury, and K.K. Biswas. A Fuzzy Theoretic Approach for Video Segmentation Using Syntactic Features. *To appear in Nov 2001, Pattern Recognition Letters*, 22:1359–1369, Nov 2001.

[9] Changick Kim and Jenq-Neng Hwang. An Integrated Scheme for Object-Based Video Abstraction. *ACM Electronic Proceeding*, http://www.acm.org/sigs/sigmm/MM2000/ep/kim, 2000.

[10] T.P. Minka and R.W. picard. Interactive Learning Using a Society of Models. In *Proc. CVPR, San Fransisco*, June 1996.

[11] R. Ruiloba, P. Joly, Marchand-Millet S, and G. Quenot. Towards a Standard Protocol for the Evaluation of Temporal Video Segmentation Algorithms. In *European Workshop on Content-Based Multimedia Indexing, CBMI'99*, October 25-27 1999.

[12] Yuhui Shi, Russell Eberhart, and Yaobin Chen. Implementation of Evolutionary Fuzzy Systems. *IEEE Transactions on Fuzzy Systems*, 7(2):109–118, April 1999.

[13] G. Sudhir, J. Lee, and A.K. Jain. Automatic Classifiaction of Tennis Video for High-level Content-based Retrieval. In *Proc. IEEE International Workshop on CAIVD, Bombay, India*, 3 Jan 1998.

[14] X. Sun, M.S. Kankanhali, Y. Zhu, and J. Wu. Content-Based Representative Frame Extraction for Digital Video. In *IEEE Conference of Multimedia Computing and Systems, Austin TX*, 1998.

[15] M. Szummer and R. Picard. Indoor-outdoor Image Classification. In *Proc. IEEE International Workshop on CAIVD, Bombay, India*, 3 Jan 1998.

[16] Y. Taniguchi, A. Akutsu, and Y. Tonomura. PanoramaExcerpts: Extracting and Packing Panoromas for Video Browsing. In *Proc. ACM Mutimedia 97, Seattle, WA*, pages 427–436, 1997.

[17] Y. Tonomura, A. Akutsu, K. Otsuji, and T. Sadakata. Videomap and Videospaceicon: Tools for Anatomizing Video Content. In *Proc. ACM INTERCHI'93*, pages 131–141, 1993.

[18] H. Ueda, T. Miyatake, and S. Yoshizawa. An Interactive Natural-Motion-Picture Dedicated Multimedia Authoring System. In *Proc. ACM SIGCHI91, New Orleans, LA*, pages 343–350, 1991.

[19] M.M. Yeung B.L. Yeo, W. Wolf, , and B. Liu. Video browsing using clustering and scene transtions on compressed sequences. In *Proc. of IS&T/SPIE Multimedia Computing and Networking*, 1995.

[20] H.J. Zhang, Yihong Gong, S.W. Smoliar, and Shuang Yeo Tan. Automatic parsing of news video. In *Proc. Of Int. Conf. On Multimedia Computing and Systems, Boston, Massachusetts, USA*, pages 45–54, May 1994.

[21] H.J. Zhang, C.Y. Low, S.W. Smoliar, and J.H. Wu. Video Parsing Retrieval and Browsing : an Integrated and Content Based Solution. In *Proc. of Multimedia '95, San Francisco, CA USA*, pages 15–24, 1995.

[22] Wensheng Zhou, Asha Vellaikal, and C.-C. Jay Kuo. Rule-based Video Classification System for Basketball Video Indexing. In *Proc. ACM Multimedia Workshop Marina Del Rey CA USA*, pages 213–216, 2000.