Sport Video Shot Segmentation and Classification

Rozenn Dahyot, Niall Rea and Anil Kokaram

Electronic and Electrical Engineering Department
Trinity College, Dublin 2, IRELAND
E-mail: dahyot@mee.tcd.ie, {akokaram,oriabhan}@tcd.ie
http://www.mee.tcd.ie/~ sigmedia

ABSTRACT

This paper considers the statistics of local appearance based measures that are suitable for the visual parsing of sport events. The moments of the colour information are computed, and the shape content in the frames is characterised by the moments of local shape measures. Their generation process is very low cost. The temporal evolution of the features then is modelled with a Hidden Markov Model. The HMM is used to generate higher level information by classifying the shots as close ups, court views, crowd shots and so on. The paper illustrates how those simple features, coupled with the HMM, can be used for parsing snooker and tennis footages.

Keywords: Statistical Moments, Hidden Markov Model, shot classification

1. INTRODUCTION

Retrieval and summarisation of sports footage has received increasing interest in recent years. The growing amount of video data available in the home has made it attractive to consider simplifying user interactivity with the potential "home sports media base". Automated retrieval from any video media typically involves segmenting the media temporally into shot segments and then attempting to uncover patterns in the temporal evolution of the shots. Both of these problems are difficult in general because the nature of the data is not well understood. The important aspect about sports media is that it has a well-defined context both in terms of the organisation of the footage and the semantics of the game itself. Therefore it is conceivable that for sports at least, it is possible to create algorithms that in some sense understand the evolution of this type of media and so allow more intimate user interactivity.[1–4]

To extract highlights from sport events, shot changes are often detected using the probability density functions of local appearance based measures of each frame. Colour histograms, for example, approximate the distribution of the colours, can be used to perform shot cut detection. Temporal changes are detected by comparing histograms computed on two successive frames.[5,6] These features then allow shots to be classified. All meaningful events related to the state of play of the game are captured in the full view of the playing area. For instance, shots displaying the complete snooker table in snooker broadcast footage can be detected using the moment of the Hough transform of the edges which had been segmented by performing a colour thresholding operation.[3] In this view balls can be tracked and pots can be detected. Since no camera motion occurs in those views of the entire snooker table, the moments remain constant over the images of those shots. The same method has been successfully used to detect rallies in tennis broadcast.[4] The value of this method is that, unlike previous work, there is no need to create 3D models of the scene.[7] However, the method is sensitive to the choice of the threshold used to segment the image.

The paper extends the idea of using statistical moments further by first of all acknowledging that the colour information can characterise the playing surface in different types of sports events e.g. green for snooker, green/blue/orange etc. for tennis depending on the tournament, etc. In effect, statistics of chrominance information remain constant when the images contain a large proportion of the pixels of the playground. Therefore, using this property, shots of the playing area can be easily detected. The playing area is also characterised by strong and straight edges due to its borders or the delineating field lines.[3] Using the first derivatives of the images, several local-appearance based features can be computed to describe the shape of the playing area. Three measures, relevant for characterising straight edges,[8] have been proposed: the norm of the first derivatives, the angle and a measure of the alignment. The 3-D histogram computed using this shape feature, is similar to the Hough transform,[8] and its moments, presented in section 2, are used in this paper to classify shots.

Visual Communications and Image Processing 2003, Touradj Ebrahimi, Thomas Sikora, Editors, Proceedings of SPIE Vol. 5150 (2003) © 2003 SPIE · 0277-786X/03/$15.00

Shot classification is possible once some reasonable parsing of the sequence has been obtained. Hidden Markov models (HMM) have been used successfully in video indexing problems such as transition detection and sccene classification[9] or highlights extraction in sport videos.[2] In section 3, we propose to use Hidden Markov model to perform the classification of shots using these new features.[10] The temporal evolution of the feature values is modelled using a two state HMM for each particular camera view. There are five possible views in snooker footage, and three in tennis footage. A maximum likelihood approach is used to select the best matching view model given the observed feature data in a shot.

In section 4, experimental results show performance using footage of tennis and snooker, and in this context, a good classification performance is obtained. This work was undertaken as part of the MOUMIR (Models for Unified Multimedia Information Retrieval www.moumir.org), MUSE-DTV (Machine Understanding of Sport Events for Digital Television) and CASMS (Content Aware Sports Media Streaming) projects. These projects are supported by the EU and Enterprise Ireland respectively.

2. VISUAL FEATURES

Statistics of local appearance based measures, for instance colour information, are often used as relevant features for image and video indexing. Comparisons of histograms computed over successive frames of the sequences have been successfully performed, for instance, for shot change detection or object statistics extraction.[8]

Histograms of local appearance-based measures of images are an approximation of their probability density function. If $\mathbf{m}$ is a local appearance-based measure, the histogram computed using the set of measures $\mathbf{m}$ observed in the frame $I(t)$ is an approximation of $\mathcal{P}(\mathbf{m}|t)$, and the detection of shot changes is simply performed by measuring the (dis)similarity between $\mathcal{P}(\mathbf{m}|t)$ and $\mathcal{P}(\mathbf{m}|t+1)$. Depending on the dimension of the measure $\mathbf{m}$, computing its histogram can be computationally expensive, and can also be a bad approximation to the p.d.f. (depending on how many measures are available to compute the histogram and how many classes there are to fill in[11]).

As an alternative, the moments of the shape and colour features, corresponding to the projection of the distribution, allows the representation to be reduced.[12] The moments of the shape information in the frames have been used[3,4] to extract relevant shots in snooker and tennis video. The snooker table and the tennis court are segmented thanks to colour homogeneity. The contours of the resulting binary maps are analysed using the Hough transform. Then the Hough moments are computed and have proved to be relevant features to detect shots of the table or tennis court.[3,4]

The colour information used as a first step of the process,[3,4] is also relevant to detect shots of the playground. In addition to the shape information, we investigate the idea of using the first moment of colour information. We present the measures used to code the colour information in section 2.1. In order to avoid the computation of the Hough transform, we propose to use the shape features used in Dahyot et al.[8] These measures are reviewed in section 2.2. The computation of colour and shape moments are reviewed in section 2.3.

2.1. Colour features

Sport videos are recorded using a finite number of fixed cameras. Some of the shots, involve views of the playing area which is usually well-defined by its colour (cf. fig. 1). These particular shots are important as it is the primary view in which the main events are captured. For example in snooker[3] and in tennis.[4] The colour triplet used in this article, is defined by[8]:

$$\mathbf{m}^{color} = \begin{pmatrix} r = \frac{R}{R+G+B} \\ g = \frac{G}{R+G+B} \\ I = R+G+B \end{pmatrix} \quad (1)$$

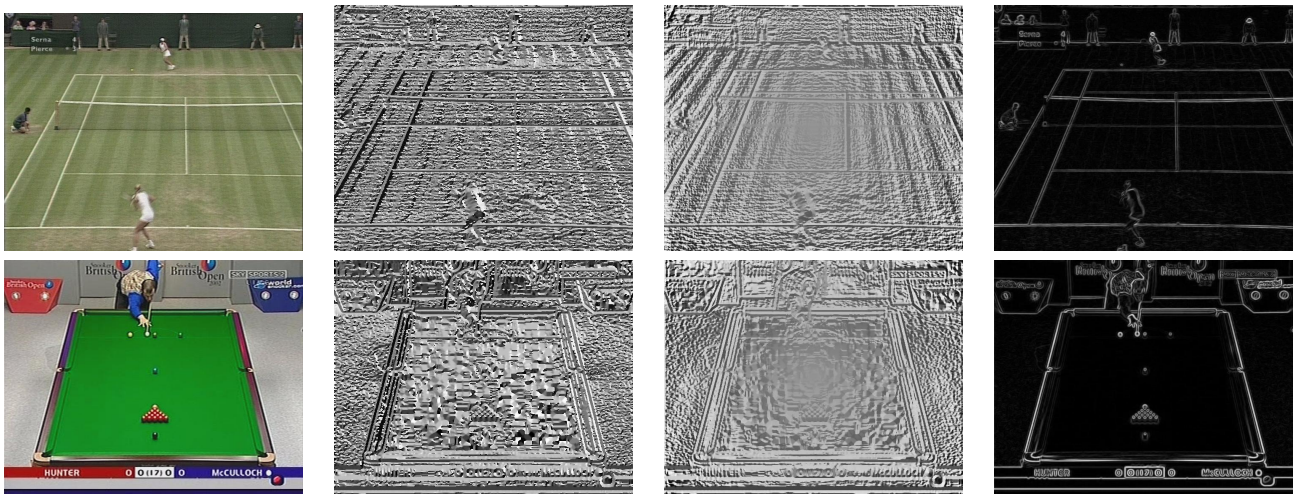
The couple (r, g) collects the chrominance information and the intensity is represented by I.

2.2. Shape features

Three parameters have been used to define a shape measurement,[8] related to the alignment of local edges. The first parameter is the angle θ of a local edge. The second , α, is an alignment measurement. Two points belonging to the same straight contour have theoretically identical (θ, α) values. The third parameter is the norm N of the gradient:

$$\mathbf{m}^{shape} = \begin{pmatrix} \theta = \arg\tan \frac{I_x}{I_y} \\ \alpha = x \, \frac{I_x}{N} + y \, \frac{I_y}{N} \\ N = \sqrt{I_x^2 + I_y^2} \end{pmatrix} \tag{2}$$

where I_x and I_y denote the two components of the spatial gradient computed over the intensity. Figure 1 shows some frames with their corresponding shape measures.



frames θ α N

Figure 1. Example of frames and their shape measures, showing the whole table in snooker video and the whole court in tennis video.

2.3. Moments

We consider the central moments of the colour $\mathcal{M}_{ijk}^{colour}$ and shape features $\mathcal{M}_{ijk}^{shape}$. The indices, (i, j, k), refer to the order on each variable respectively (r, g, I) and (θ, α, N).

First order moments $(i + j + k = 1)$. The first order moments, corresponding to the mean values of the features, are computed on each frame by:

$$\begin{aligned} \mathcal{M}_{100}^{colour}(t) &= \textstyle\sum_{\mathbf{x}} r(t, \mathbf{x}); \quad \mathcal{M}_{100}^{shape}(t) = \textstyle\sum_{\mathbf{x}} \theta(t, \mathbf{x}) \\ \mathcal{M}_{010}^{colour}(t) &= \textstyle\sum_{\mathbf{x}} g(t, \mathbf{x}); \quad \mathcal{M}_{010}^{shape}(t) = \textstyle\sum_{\mathbf{x}} \alpha(t, \mathbf{x}) \\ \mathcal{M}_{001}^{colour}(t) &= \textstyle\sum_{\mathbf{x}} I(t, \mathbf{x}); \quad \mathcal{M}_{001}^{shape}(t) = \textstyle\sum_{\mathbf{x}} N(t, \mathbf{x}) \end{aligned} \tag{3}$$

where t is the time or the number of the image in the video, and $\mathbf{x}$ is the location of the local measure in the image.

Higher order moments $i+j+k>1$**.** The features are centred on their mean values. For instance, the red component is centred as:

$$\overline{r}(t,\mathbf{x}) = r(t,\mathbf{x}) - \mathcal{M}_{100}^{colour}(t)$$

The central moments are then computed by:

$$\begin{aligned} \mathcal{M}_{ijk}^{colour}(t) &= \textstyle\sum_{\mathbf{x}} \overline{r}^i(t,\mathbf{x})\ \overline{g}^j(t,\mathbf{x})\ \overline{I}^k(t,\mathbf{x}) \\ \mathcal{M}_{ijk}^{shape}(t) &= \textstyle\sum_{\mathbf{x}} \overline{\theta}^i(t,\mathbf{x})\ \overline{\alpha}^j(t,\mathbf{x})\ \overline{N}^k(t,\mathbf{x}) \end{aligned} \tag{4}$$

2.4. The relevance of the moments

Colour information. We are mainly interested in shots showing the playing area. Specifically the whole table view in snooker videos, or the whole court in tennis (cf. images fig. 1). These shots gather the main information concerning the game itself. Figure 2 shows the first colour moments : each dot corresponds to the couple of the chrominance information $(\mathcal{M}_{100}^{colour}, \mathcal{M}_{010}^{colour})$ measured for each image over the tennis sequence *Pierce* and the sequence *snooker3* (cf. table 1: the colour of the dots indicates to which kind of shot the frame belongs to). The green (or light grey pointed with the arrow) colour indicates the frames extracted from shots of the whole court/table. We can note that those shots of interest, have relatively constant moments and form a cluster easily separable from the others. The intensity information, a priori sensitive to shadow effects, has not been used.

Shape information. The shots showing the view of the whole table in snooker videos and the whole tennis court in tennis footage, show strong straight contours delimiting the playground area, as underlined by images of the measures N in figure 1. Figure 3 shows the results of the method[3] using the hough transform of the segmented playground, and the second moment of the norm $\mathcal{M}_{002}^{shape}(t)$. Intuitively, this moment seems very relevant to detect the shots showing a large view of the snooker table: high values (in green/light grey) of $\mathcal{M}_{002}^{shape}(t)$ correspond to the view of the whole snooker table. This observation has been noticed for both snooker videos *Snooker3* and *Hunter*.

The angle (and so the alignment) is not relevant on uniform areas. In fact, assuming a Gaussian noise over the frame, the distribution of the measure θ is uniform on non-textured regions of the image.[13] This property can affect the significance of the moments of those features. The first way to deal with this problem is to segment the contours of the frames, and to consider only their local shape measures.[3] This implies the choice of thresholds that may vary depending on the quality of the footage or the recording conditions. To avoid the strict segmentation of the edges, the local shape measures can also be weighted according to their relevance.[14] The value of the norm is relevant to define weights for the angle and the alignments[13] and some shape moments, as $\mathcal{M}_{101}^{shape}$ or $\mathcal{M}_{011}^{shape}$, can be interpreted as weighted statistics of both measures θ and α. Several experiments have been performed (cf. section 4) in order to select the most relevant moments for shot classification.

3. SHOT CLASSIFICATION USING HMM

As shown in Figure 3, the evolution of the feature vectors is closely related to the view in the image. If the temporal behaviour of the feature vector for each camera view can be modelled, it is then possible to categorise each shot. The Hidden Markov Model (HMM) allows a rich variety of temporal behaviours to be modelled. This approach has been used to good effect in cognition based systems.[15,16] A two state ergodic HMM was found empirically to sufficiently model the feature vector for each of the camera view models. In order to generate an alphabet for the discrete HMM the features were quantised using a Gaussian Mixture model with parameters estimated using the EM algorithm. The mixture was pruned appropriately to reduce the size of the codebook and each feature attached to its appropriate cluster.

A Gaussian Mixture model[17] was found to be much more reliable than the K-means algorithm (typically used) since it is clear from fig. 2 that the clusters of feature points can not be modelled by a single gaussian distribution. Therefore, the Gaussian Mixture model implicitly assigns several Gaussians to model each such cluster 2. This could lead to an overquantisation of the feature space which would imply some detrimental effect

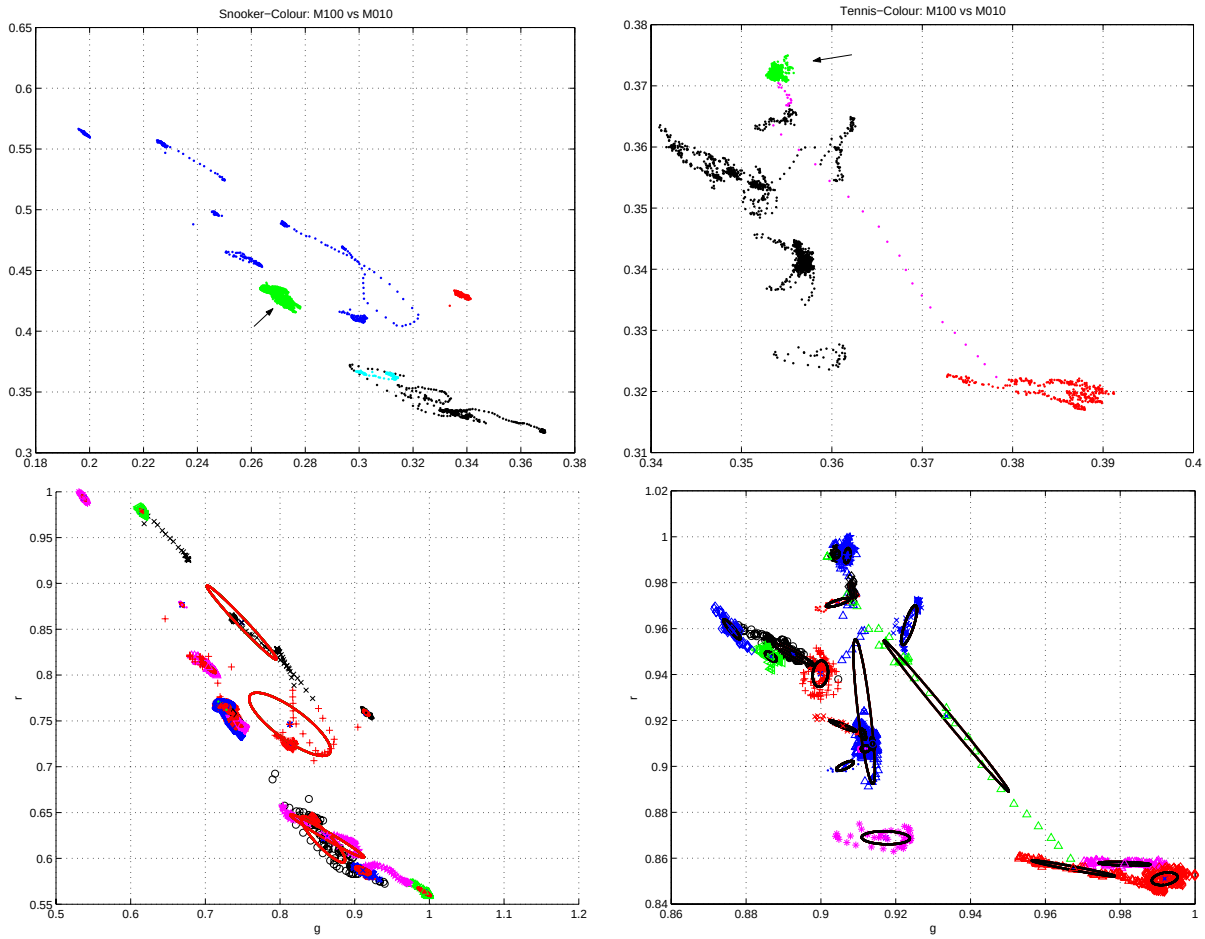


Figure 2. Colour moments distribution (Top) $(\mathcal{M}_{100}^{colour}, \mathcal{M}_{010}^{colour})$ (red={crowd, journalist}, green={whole court or whole snooker table}, blue={corners of the snooker table}, black={close − up of the players}). (Bottom) Clustered moments using a GMM, 22 mixtures for snooker and 20 for tennis

on the HMM training process. To overcome this problem, the Gaussians estimated by the EM Mixture modelling process are merged and pruned. Merging is done by applying a maximum euclidean distance threshold between the centres of the Gaussians. A distance of 0.01 is used here. Pruning is achieved by eliminating mixtures whos probability or determinant of the covariance matrix is zero. If the conditional number of the covariance matrix exceeds a specified threshold $(10,000)$ that mixture is also eliminated. For the colour moments for the snooker footage for example, 25 initial mixtures were used to model the distribution. This was pruned to 22 mixtures and merged to 14. Each of these clusters is then assigned a discrete codebook entry for training and classification via the HMM.

Knowing the number of states and discrete codebook entries, a model λ, can be defined for each of the competing camera views. A succinct definition of a HMM is given by the following parameters, where K is the number of classes, N is the number of states and M is the number of unique observation symbols per state.

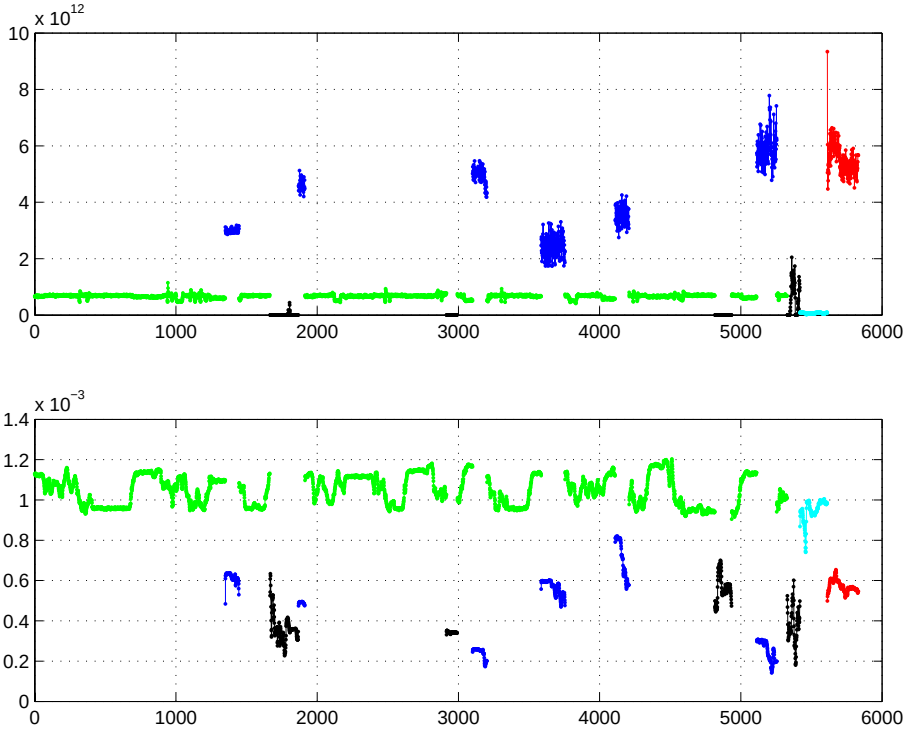


Figure 3. Comparison of shape moments : Top, the Hough moment used in Denman et al.[3]; Bottom, the variance of the norm $\mathcal{M}_{002}^{shape}(t)$.

$$
\begin{aligned}
A &= \{a_{ij}\} = P(q_{t+1} = S_j | q_t = S_i), \;\; 1 \leq i, j \leq N \\
B &= \{b_j(O_t)\} = P(O_t = v_k | q_t = S_j), \;\; 1 \leq j \leq N, 1 \leq k \leq M \\
\pi &= \{\pi_i\} = P(q_1 = S_i), \;\; 1 \leq i \leq N \\
\lambda_K &= (A, B, \pi)
\end{aligned}
\tag{5}
$$

The parameters are defined as follows : A is a state transition probability matrix, B is an observation probability matrix or confusion matrix, in the discrete case, and π is a vector of initial state probabilities. An extensive tutorial on HMMs is available from Rabiner.[18]

The parameter reestimation process was conducted by training the system with the ground truth. Half of the feature vectors from each camera view were used to train each model. The Baum-Welch algorithm was then used to find the maximum likelihood model parameters that best fit the training data. The models for each source were then tested against each sequence individually ($\approx$ 4 minutes of *Snooker3* snooker footage, 2 minutes of *Pierce* tennis footage and 16 minutes of *Hunter* snooker footage). The shot cuts were detected offline using a combination of the gradient of a fourth order moment of Hough transform[3] of the image and the sum of histogram differences of the luminance component of the image sequence. Given this observation sequence the probability $P(O|\lambda)$ can be calculated. Each camera view can then be classified by finding the model that results in the greatest likelihood of occurring according to:

$$C = \arg \max_{1 \leq x \leq K} [P(O|\lambda_x)] \tag{6}$$

The whole process is summarized in figure 4.

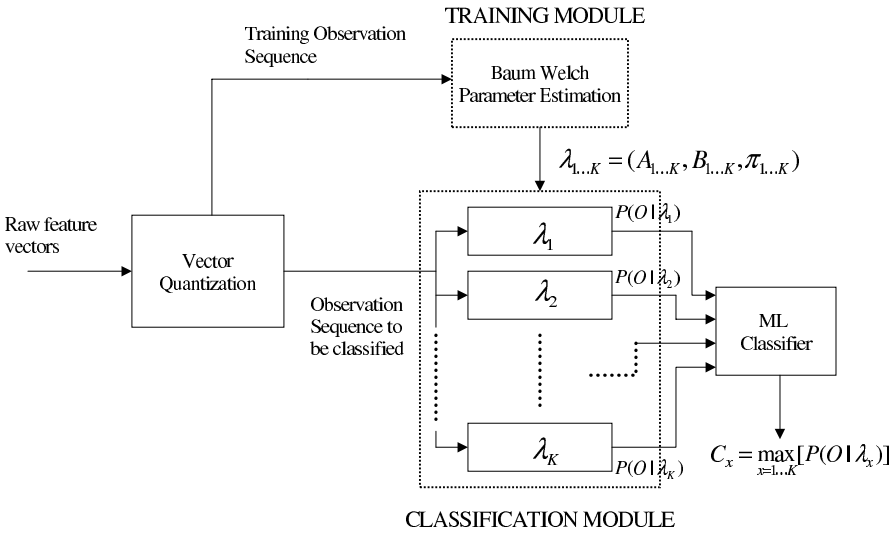


Figure 4. ML-classifier.

4. EXPERIMENTAL RESULTS

Sections 4.1 and 4.2 present the results of classification of shots using colour and shape moments. Experiments in section 4.1 are performed using different combination of moments over the two short sequences called *Snooker3* and *Pierce* (cf. table 1). Section 4.2 presents the results of classification over the longer snooker video *Hunter*. Table 1 indicates the length of each sequence (the number of frames), the number of shots to classify and the chosen number of classes of interest. Some frames of those classes are shown in figure 5.

	Snooker3 (snooker)	*Pierce* (tennis)	*Hunter* (snooker)
nb frames	5833	2949	24 251
nb shots	21	16	115
nb classes	5	3	5

Table 1. Sequences used in the experiments.

Assessment of the results. The results are assessed by computing the recall and the precision of the system[9, 19]:

$$Recall = \frac{A}{A+C} \quad Precision = \frac{A}{A+B} \tag{7}$$

with A the correctly retrieved, B the falsely retrieved and C the missed.

Merging the performances. The system is initially assessed separately using colour and shape information. Small dimensional measures such as a couple of moments, allow memory to be saved in the indexing system, and also are easy and fast to compute. We restrict attention to two dimensional combinations of features in the experiments. The classification system is therefore assessed using a couple of colour moments and some couples of shape moments. Their results are complementary, so merging them enables an improvement in the classification. Several merging schemes have been tried. For instance, considering the two kind of observations, shape and colour, independently , the merged decision can be performed by:

$$C = \arg \max_{1 \leq x \leq K} \left[P(O^{colour}|\lambda_x^{colour}) \; P(O^{shape}|\lambda_x^{shape}) \right]$$

Figure 5. Example of frames in our classes of interest in the footages: from top to bottom, *Snooker3, Pierce, Hunter.*

Nevertheless, in the following experiments the best improvement in merging has been obtained by considering as a first result the classification by colour moments. Then, the unclassified shots using colour information, are classified using shape information.

4.1. Experiments on short sequences

Experiment 1: Shot classification in Snooker video. The snooker sequence *snooker3* is considered for this experiment. Using the two colour moments $(\mathcal{M}_{100}^{colour}, \mathcal{M}_{010}^{colour})$, the system succeeds in correctly classifying the shots with 95% recall and 100% precision.

Using only the variance of the norm $\mathcal{M}_{002}^{shape}$, the system classifies shots with 48% recall and 100% precision. Table 2 collects the results of the classification over the sequence *snooker3* using several combination of shape moments with $\mathcal{M}_{002}^{shape}$ always used (cf. section 2.4).

Moments	$(\mathcal{M}_{002}^{shape}, \mathcal{M}_{100}^{shape})$	$(\mathcal{M}_{002}^{shape}, \mathcal{M}_{010}^{shape})$	$(\mathcal{M}_{002}^{shape}, \mathcal{M}_{001}^{shape})$	$(\mathcal{M}_{002}^{shape}, \mathcal{M}_{011}^{shape})$	$(\mathcal{M}_{002}^{shape}, \mathcal{M}_{101}^{shape})$
Recall	60%	72%	67%	70%	48%
Precision	92%	81%	100%	93%	100%

Table 2. Results of the classification using different combination of shape moments on the video *Snooker3*.

While the results using the colour moments are already very good (only one shot out of 21 is missed), shape moments lead to an unremarkable classification performance. Using $\mathcal{M}_{101}^{shape}$ does not improve the result obtained by $\mathcal{M}_{002}^{shape}$ used alone. However, the addition of $\mathcal{M}_{010}^{shape}$ or $\mathcal{M}_{011}^{shape}$ improves the results of the classification. By combining colour and shape information $(\mathcal{M}_{100}^{colour}, \mathcal{M}_{010}^{colour})$ and $(\mathcal{M}_{002}^{shape}, \mathcal{M}_{010}^{shape})$, in the manner discussed previously, yields a classification with 100% recall and precision.

Experiment 2: Shot classification in Tennis video. The *Pierce* sequence is considered for this experiment. The colour moments $(\mathcal{M}_{100}^{colour}, \mathcal{M}_{010}^{colour})$ yield classification with 56% recall and 100% precision. Excluding the couples $(\mathcal{M}_{002}^{shape}, \mathcal{M}_{011}^{shape})$ and $(\mathcal{M}_{002}^{shape}, \mathcal{M}_{010}^{shape})$, shape moments do not perform better, as shown in table 3. Nevertheless, merging shape $(\mathcal{M}_{100}^{colour}, \mathcal{M}_{010}^{colour})$ and colour information allows classification of the shots of the *Pierce* sequence with an encouraging 87% recall and 93% precision.

Moments	$(\mathcal{M}_{002}^{shape}, \mathcal{M}_{100}^{shape})$	$(\mathcal{M}_{002}^{shape}, \mathcal{M}_{010}^{shape})$	$(\mathcal{M}_{002}^{shape}, \mathcal{M}_{001}^{shape})$	$(\mathcal{M}_{002}^{shape}, \mathcal{M}_{011}^{shape})$	$(\mathcal{M}_{002}^{shape}, \mathcal{M}_{101}^{shape})$
Recall	53%	60%	53%	84%	50%
Precision	89%	90%	89%	84%	100%

Table 3. Results of the classification using different of moments in the video *pierce*.

Some Remarks. These first classification results are encouraging, especially when the performances of both colour and shape information are merged. Some mis-classifications appear mainly because the training of the HMM process suffers from only considering half of the shots in the short sequences *Pierce* and *Snooker3*. Hence, some events occurring at the end of the shots, such as camera motion for instance, are not learned and generate some unpredictable variations over the features for our trained classification system. Section 4.2 proposes to assess the system on the longer sequence *Hunter*.

4.2. Application on a long sequence

The colour moments $(\mathcal{M}_{100}^{colour}, \mathcal{M}_{010}^{colour})$ yield 79% recall and 84% precision. In this experiment, similar results are obtained using several combination of the shape moments (cf. table 4). As in the previous experiments,

Moments	$(\mathcal{M}_{002}^{shape}, \mathcal{M}_{100}^{shape})$	$(\mathcal{M}_{002}^{shape}, \mathcal{M}_{010}^{shape})$	$(\mathcal{M}_{002}^{shape}, \mathcal{M}_{001}^{shape})$	$(\mathcal{M}_{002}^{shape}, \mathcal{M}_{011}^{shape})$	$(\mathcal{M}_{002}^{shape}, \mathcal{M}_{101}^{shape})$
Recall	99%	95%	95%	97%	95%
Precision	71%	79%	81%	79%	70%

Table 4. Results of the classification using different of moments in the video *Hunter*.

merging both colour and shape information $(\mathcal{M}_{002}^{shape}, \mathcal{M}_{010}^{shape})$ inproves the performance to 100% recall and 84% precision.

5. CONCLUSION AND FUTURE WORK

In this article, we have proposed a method for shot classification using HMM where moments of local appearance based measures, computed over each frame of the shot, have been used as observations. Experiments have been carried out on sport video sequences, snooker and tennis, and our classification scheme has shown good performance. Further possible improvements will be investigated using other (orthogonal) moments,[12] and involving the audio content of the videos.[4, 9]

REFERENCES

1. Y. Rui, A. Gupta, and A. Acero, "Automatically extracting highlights for tv baseball programs," in *proceedings of ACM Multimedia Conference*, (Los Angeles, California), October 2000.
2. P. Chang, M. Han, and Y. Gong, "Highlight detection and classification of baseball game video with hidden markov models," in *International Conference on Image Processing*, 2002.
3. H. Denman, N. Rea, and A. Kokaram, "Content based analysis for video from snooker broadcasts," in *proceedings of International Conference on Image and Video Retrieval (CIVR)*, (London,UK), July 2002.
4. R. Dahyot, A. C. Kokaram, N. Rea, and H. Denman, "Joint audio visual retrieval for tennis broadcasts," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, (Hong Kong, CHINA), April 2003.
5. J. S. Boreczky and L. A. Rowe, "Comparison of video shot boundary detection techniques," in *SPIE proceedings of Storage and Retrieval for Image and Video Databases*, pp. 170–179, 1996.
6. J. Yu, G. Bozdagi, and S. Harrington, "Feature-based hierarchical video segmentation," in *International Conference on Image Processing*, **2**, (Washington, DC), October, 26-29 1997.

7. F. Sudhir, J. Lee, and A. Jain, "Automatic classification of tennis video for high-level content-based retrieval," in *proceedings of International workshop on Content-Based Access of Image and Video Databases (CAIVD'98)*, 1998.
8. R. Dahyot, P. Charbonnier, and F. Heitz, "Unsupervised statistical change detection in camera-in-motion video," in *IEEE proceedings of the International Conference on Image Processing*, (Thessaloniki, GREECE), October 2001.
9. J. S. Boreczky and L. D. Wilcox, "A hidden markov model framework for video segmentation using audio and image features," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, **6**, pp. 3741–3744, (Seattle, WA), 1998.
10. J. Assfalg, M. Bertini, A. D. Bimbo, W. Nunziati, and P. Pala, "Soccer highlight detection and recognition using hmms," in *IEEE International Conference on Multimedia and Expo*, 2002.
11. C. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1995.
12. M. K. Mandal, T. Aboulnasr, and S. Panchanathan, "Image indexing using moments and wavelets," *IEEE Transactions on Consumer Electronics* **42**, pp. 557–565, August 1996.
13. R. Dahyot, *Appearance based road scene video analysis for the management of the road network*. PhD thesis, Louis Pasteur University, Strasbourg, November 2001.
14. C. Vertan and N. Boujemaa, "Spatially constrained color distribution for image indexing," in *proceedings of International Conference on Color in Graphics and Image Processing (CGIP)*, (Saint Etienne, France), October 2000.
15. T. H. I. Cohen, A. Garg, "Emotion recognition from facial expressions using multilevel hmm," in *Neural Information Processing Systems*, 2000.
16. W. J. M. Petkovic, "Content-based video retrieval by integrating spatio-temporal and stochastic recognition of events," in *IEEE Workshop on Detection and Recognition of Events in Video*, 2001.
17. P. K. I. F. Sbalzariniy, J. Theriot, "Machine learning for biological trajectory classifcation applications," in *Center for Turbulance Research. Summer Program*, 2002.
18. B. J. L.R. Rabiner, "An introduction to hidden markov models," *IEEE ASSP Mag.* , pp. 4–16, 1986.
19. A. D. Bimbo, *Visual Information Retrieval*, Academic Press Morgan Kaufmann Publishers, 1999.