HMM Based Classification of Sports Videos Using Color Feature

Josh Hanna*, Fatma Patlar†, Akhan Akbulut†, Engin Mendi* and Coskun Bayrak*†

*Computer Science Department, University of Arkansas at Little Rock, AR, USA

{jjhanna | esmendi | cxbayrak}@ualr.edu

†Computer Engineering Department, Istanbul Kultur University, Istanbul, Turkey

{f.patlar | a.akbulut | c.bayrak}@iku.edu.tr

***Abstract*—Video content classification is an important element for efficient access and retrieval of video in any media content management system. Categorizing the video segments can help to provide convenience and ease in accessing the relevant video content without sequential scanning. In this paper, we present a Hidden Markov Model (HMM) based classification technique for sports videos. Speed of color changes is computed for each video frame and used as observation sequences in HMM for classification. Experiments using more than 1 hour of 18 training and 18 testing sports videos of 3 predefined genres (golf, hockey and football) give very satisfactory classification accuracy.**

I. Introduction

Multimedia content classification refers to the computerized apprehension of the semantic meanings of a multimedia file or document. With the increase in digital video contents, efficient techniques for classification of videos according to their contents have become more important. Applications such as digital libraries, e-Learning, video-on-demand, digital video broadcast and interactive TV generate and use large collections of video data. For an effective use of these video data, all digital contents must be classified based on their categories. There has been a growing demand for content based automatic video classification for the web multimedia administration. For this reason, numerous research is being done for such systems.

Several content based classification systems for organizing and managing video databases have been recently proposed. Classification of the videos into predefined genres is the most prefered. Basic working principle for this type of applications is classical pattern classification algorithm [1]. First, features like color, sound or video text are extracted from the videos, then passed from a reduction process to be ready for the classification.

In [2], nearest neighbor clustering is used for video classification. A more complex framework is represented as fully automatic and computationally efficient framework for analysis and summarization of soccer videos using cinematic and object-based features. This model uses cinematic and object-based features for semantic analysis of sports videos [3].

Extracted features are commonly classified with HMM for segmenting video contents. Boreczky and Lynn [4] used three types of features for video segmentation; the standard histogram difference, an audio distance measure and an estimate of object motion between two adjacent frames. Other implementations operated object color and texture features to generate highlights for soccer videos [5]. Zhu [6] classified news stories using features obtained from closed captions. This work is an example for video classification using only text features. Liu [7] [8] [9] used audio features such as non-silence ratio, volume standard deviation, volume dynamic range, pitch standard deviation, voice/music ratio, noise/unvoice ratio, frequency centroid and frequency bandwidth. Those features are extracted from the segments of the sampled audio signals and used in one-class-one-network structure for classification.

In this paper, we present a video classification approach based on HMM for video content classification using color feature. Our aim is to categorize the input video from the predefined groups: golf, hockey and football. The rest of this paper is organized as follows. Section 2 presents the concept of applying HMM for video classification and our feature extraction details. Experimental results and conclusion are given in Section 3 and Section 4, respectively.

II. HMM for Video Classification

HMM is a popular technique widely used in signal processing. HMMs are a formal foundation for making probabilistic models of linear sequence "labeling" problems [10] and they are especially known for their applications in temporal pattern recognition such as speech, handwriting, gesture recognition, part-of-speech tagging, musical score following, partial discharges and bioinformatics. They are mostly used for classifying sequential data to capture the temporal relationships of the extracted features. In our research, we extended it to video analysis and classification.

A. *Definition of HMM*

In an HMM, there are a finite number of states, each of which is associated with a transition probability to the others. Everytime, the HMM stays in one definite state. The states at time t is directly influenced by the state at time $t-1$. After each translation from one state to another, an output observation is generated based on an observation probability distribution associated with the current states [10]. Formally, a HMM is defined to be:

$$HMM = \{N, B, \Pi\}$$

where N is the set of states, B is the number of observation symbols and Π is set of state transition probabilities.

B. Feature Extraction

Each video is considered to be a sequence of images with each image being represented with a color feature. Color data from each pixel in RGB color space is gathered and averaged for each frame. Using red, green and blue saturations, the speed of color change is calculated by subtracting each color's saturation from the saturation of the previous frame. Let $R_i(\gamma)$, $G_i(\gamma)$ and $B_i(\gamma)$ be the red, green and blue channels associated with the pixel at position γ in frame i, respectively. The speed of red, green and blue channels S_R, S_G and S_B in frame i is calculated by:

$$
\begin{aligned}
S_R &= \sum_{\gamma} \frac{[R_i(\gamma) - R_{i-1}(\gamma)]^2}{R_i^2(\gamma)} \\
S_G &= \sum_{\gamma} \frac{[G_i(\gamma) - G_{i-1}(\gamma)]^2}{G_i^2(\gamma)} \\
S_B &= \sum_{\gamma} \frac{[B_i(\gamma) - B_{i-1}(\gamma)]^2}{B_i^2(\gamma)}
\end{aligned}
\tag{1}
$$

C. HMM Classification

An embedded HMM is used which generalized in Embedded Bayesian Networks for the classification studies. As illustrated in Figure 1, using conditional independences between the feature vectors in the HMM structure enable to convert 3D state transition matrix to 2-dimensional vertical and horizontal state transition matrices.

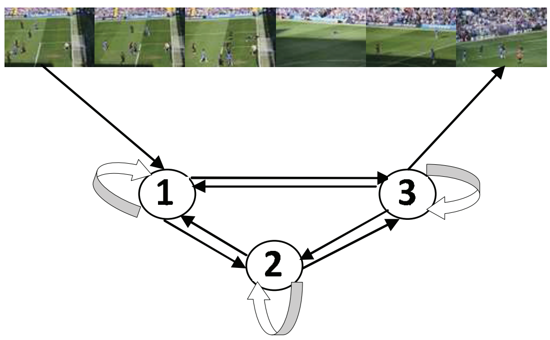


Fig. 1: Sequential representation of the 3 state HMM.

Figure 2 shows the HMM classification scheme. Process starts with a given unknown-genre video. First, an observation sequence based on the speed of color changes computed from video frames is generated from this input video. Then, the observation sequences are fed into each HMM. Finally, computing the log-likelihood of the test sequences, incoming videos are classified.

In every step, the system determines the probability distribution of the next phase from the given state. Classification process of videos consists of two step. In the first step Baum-Welch algorithm is used to estimate the most likely parameters for the HMM that generates the training set. In the second step the probability of the observation of that sequence given the HMM is compute by backward algorithm. Property groups identified as using a similar color distributions of the groups were calculated with Gaussian mixture.

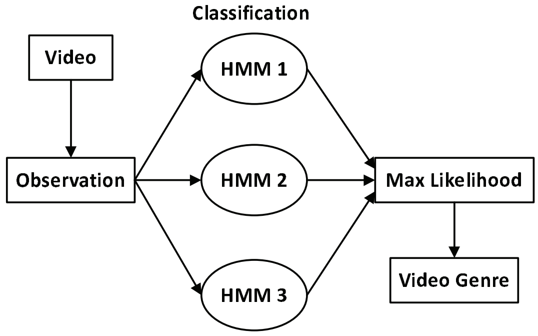


Fig. 2: HMM classification scheme.

III. Experimental Results

The system was developed with Java and Matlab. Feature extraction part was implemented in Java; HMM-based classification part was implemented in Matlab. We evaluated our system by classifying sports videos into golf, hockey and football. The experiments were carried out on more than 1 hour of sports video data. Table 1 shows the number of data set we used. The whole data set contains 18 training videos: 6 for golf, 6 for hockey and 6 for football. 18 videos (6 for golf, 6 for hockey and 6 for football) were used as a testing set.

TABLE I: Data set

	Golf	Hockey	Football
Training Set	6	6	6
Test Set	6	6	6

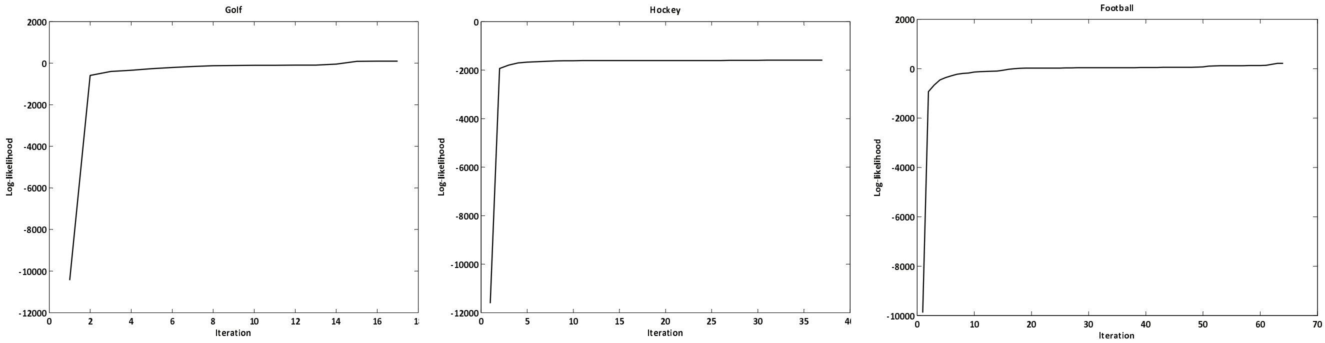
Learning curves of the model are given in Figure 3. The figure indicates the training stops when the log-probability has stabilized at which no additional learning is needed. As shown in the figure, the learning curves of sampled categories have almost flattened in about 5 iterations.

Figure 4 shows the log likelihood of the HMM outputs from test video sequences. The plots present the probability of proposed HMM-based model on sampled categories.

To evaluate the performance of the proposed framework, we computed precision and recall as follows:

$$
\begin{aligned}
\text{Precision} &= \frac{\text{true positive}}{\text{true positive} + \text{false positive}} \\
\text{Recall} &= \frac{\text{true positive}}{\text{true positive + false negative}}
\end{aligned}
\tag{2}
$$

Table 2 depicts the classification results with various states. As shown in the table, for the state number of 4, 100% classification accuracy is obtained.



(a) (b) (c)

Fig. 3: Likelihood learning curve of HMM model throughout training, a) golf, b) hockey, and c) football.

TABLE II: Classification results (%)

	Golf		Hockey		Football	
state number	Precision	Recall	Precision	Recall	Precision	Recall
2	75	100	100	100	100	84
3	75	100	100	100	100	67
4	100	100	100	100	100	100

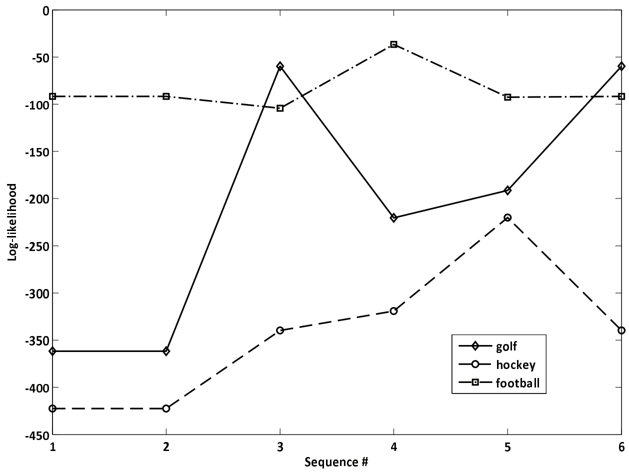


Fig. 4: The log likelihood of the HMM outputs.

IV. Conclusion

In multimedia information processing, automated video genre classification is an important problem with many applications. Categorizing video sequences can help to reduce the growing amount of cluttered video data. In this paper, we presented a video classification scheme based on HMM. Speed of color changes computed from video frames are fed to HMM as observation sequences. We tested our system on more than 1 hour of 18 training and 18 testing sports videos of 3 predefined categories: golf, hockey and football, and our results showed a very promising correct classification rate.

References

[1] A. Hauptmann and M. Witbrock, "Story segmentation and detection of commercials in broadcast news video," in *Research and Technology Advances in Digital Libraries, 1998. ADL 98. Proceedings. IEEE International Forum on*, apr 1998, pp. 168–179.

[2] W. Zhou, A. Vellaikal, and C. C. J. Kuo, "Rule-based video classification system for basketball video indexing," in *Proceedings of the 2000 ACM workshops on Multimedia*, ser. MULTIMEDIA '00, 2000, pp. 213–216.

[3] A. Ekin, A. Tekalp, and R. Mehrotra, "Automatic soccer video analysis and summarization," *Image Processing, IEEE Transactions on*, vol. 12, no. 7, pp. 796–807, july 2003.

[4] J. Boreczky and L. Wilcox, "A hidden markov model framework for video segmentation using audio and image features," in *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, may 1998, pp. 3741 –3744 vol.6.

[5] D. Yow, B. lock Yeo, M. Yeung, and B. Liu, "Analysis and presentation of soccer highlights from digital video," in *Proc. ACCV*, 1995, pp. 499–503.

[6] W. Zhu, C. Toklu, and S.-P. Liou, "Automatic news video segmentation and categorization based on closed-captioned text," in *Multimedia and Expo, 2001. ICME 2001. IEEE International Conference on*, aug. 2001, pp. 829–832.

[7] Z. Liu, Y. Wang, and T. Chen, "Audio feature extraction and analysis for scene segmentation and classification," in *Journal of VLSI Signal Processing System*, 1998, pp. 61–79.

[8] Z. Liu, J. Huang, and Y. Wang, "Classification tv programs based on audio information using hidden markov model," in *Multimedia Signal Processing, 1998 IEEE Second Workshop on*, dec 1998.

[9] J. Huang, Z. Liu, Y. Wang, Y. Chen, and E. Wong, "Integration of multimodal features for video scene classification based on hmm," in *Multimedia Signal Processing, 1999 IEEE 3rd Workshop on*, 1999, pp. 53–58.

[10] L. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257 –286, feb 1989.