Highlights' recognition and learning

in soccer video by using the shots' classification and Hidden Markov Models

Rajae El Ouazzani and Rachid Oulad Haj Thami

Laboratory: SI2M, WIM
Dept.: SI3M (Système d'Information Métiers, Multimédia et Mobiles)
Ecole Nationale Supérieure d'Informatique
et d'Analyse des Systèmes (ENSIAS), B.P. 715
Mohamed V University, Souissi, Rabat, Morocco
elouazzanirajae@gmail.com, oulad@ensias.ma

RÉSUMÉ. Dans cet article, nous proposons une nouvelle méthode d'analyse des structures vidéo des matches de football. Notre travail se focalise sur la détection des moments forts comme les buts et les fautes directes en utilisant le domaine de connaissance et l'apprentissage supervisé des modèles statistiques comme les Modèles de Markov Cachés (MMCs). Notre contribution consiste à effectuer la segmentation du flux numérique en plans: loin, médium ou proche. Ensuite, le MMC subit une phase d'apprentissage sur une base de moments forts. Finalement, nous procédons à la reconnaissance de l'importance d'un segment vidéo en utilisant notre MMC. Notre système atteint une précision de 94.5%, résultat très prometteur dans la détection des événements importants.

ABSTRACT. In this article, we propose a new method for soccer video structure analysis. Our work focuses on the detection of highlights such as goals and direct free kicks by using domain knowledge and the supervised learning of statistical models such as Hidden Markov Models (HMMs). Our contribution consists of segmenting the digital stream of shots: long, medium or close-up. Then, the HMM undergoes a learning process on a database of highlights. Finally, we proceed to the recognition of the importance of a video segment by using our HMM. Our system achieves an accuracy of 94.5%, which is a very promising result in the detection of highlights.

MOTS-CLÉS : Analyse vidéo, match de football, classification des plans, Modèles de Markov Cachés, algorithme Forward-Backward, algorithme Viterbi.

KEYWORDS : Video analysis, soccer, shots' classification, Hidden Markov Models, Forward-Backward algorithm, Viterbi algorithm.

1. Introduction

The rapid and continuous evolution of video documents as well as their large volumes require the development of new methods for indexing digital streams. Our domain of interest is soccer video especially temporal sequences of high level shot states, called long, medium and close-up. The purpose of this work is to analyse continuous video streams so as to firstly have a sequence of states automatically tagged. Then, these states will be joined to have homogeneous bits classified as a semantic state (common event or highlight). Our generic analysis approach of high level structures is inspired from the temporal variations of basic video elements. These temporal variations can be captured by statistical temporal models such as HMMs. We use the shots' classification in our analysis for the following reasons : (1) a sequence of shots is very much related to the game semantic. During the game, the video producer focuses on long shots in order to keep the viewers informed of what is happening in the field. Close-up and medium shots will interrupt long shots so as to follow the player. Then, there are less long shots because no highlights take place in the whole field. Medium and close-up shots become the majority and show what is happening in a specified area in the field. (2) the shots' type can be calculated by low level features. The algorithm which we make use of is the one presented by Y. Tabii et al. [1].

Many works have been carried out in the analysis and extraction of soccer video semantics. A. Ekin et al. [2] have proposed an automatic method for soccer video analysis to extract highlights. Their study is based on the detection of goal lines followed by the detection of a game break and the appearence of the referee who is distinguished by his black uniform. Their study also uses the appearance of the three parallel lines of the penalty zone in the extraction of these highlights. Dynamic Bayesiens Networks were introduced by C.-L. Huang et al. [3] for soccer video semantic analysis. Their approach is based on the calculation of low level features. High level semantic analysis is implemented by Bayesiens Networks/ Dynamic Bayesiens Networks. In M. Bertini et al.'s work [4], engine models with finite states and an audit modeled were proposed to model highlights and to detect their occurrences automatically in two types of sports : soccer and swimming. A soccer multimodal structure by HMM is presented by E. Kijak et al. [5]. Their study is based on the extraction of the characteristics of scenes in tennis matches and on their modelisation with HMMs.

In this paper, we propose a new approach for learning soccer video highlights, such as goals, direct free kicks, penalty kicks and corner kicks. The analysis of the video begins with the segmentation of soccer video into small video highlight segments. These segments undergo a process of extraction of representative images (key images). Then, we classify them according to their type of shots (long shots' class, medium shots' class and close-up shots' class). The sequences of shots of highlights video segment are grouped in an observation set W. In our proposal, we use Hidden Markov Models in learning stochastic observation sequences given their effectiveness. Just like a Markov chain, an HMM represents a set of observation sequences in which the state of each observation is not observed, but is associeted with a probability density function (pdf). Therefore, it is a doubly stochastic process in which the observations are a random function of states and whose status changes at any moment depending on the probability of transition from the previous state. Learning with HMMs uses a finite set of states, a set of observations W, a matrix of transitions A, a matrix of observations B and a set of initial probabilities Π. The states of our model are represented by the classes of shots : long, medium and close-up states. The values of matrixes A, B and of the set Π are manually predicted by using domain knowledge. The learning process is performed by Forward-Backward and Viterbi algorithms [6, 7] which estimate final values of matrixes A and B and of the set Π. After the learning phase, we undertake a process to recognize the highlights. This phase uses the two algorithms mentioned above (Forward-Backward and Viterbi) to calculate a probability whose value determines the importance of the entry observation sequence of our HMM. The value of the probability is then compared to a predefined threshold.

This paper is organised as follows. In Section 2, we present an analysis of the video program. In Section 3, we modelize the highlights by using HMMs. In Section 4, we propose the learning and recognition algorithms in HMMs. In Section 5, we present the experimental results. In Section 6, we compare our work with others found in the literature. Finally, in Section 7, we give a conclusion.

2. Video program analysis

In the literature, various approches have been proposed to recognize the highlights in soccer video. For instance, Dynamic Bayesiens Networks (DBN) have been made use of by C.-L. Huang et al. [3], and the appearance of the referee is used to record highlights in soccer video by A. Ekin et al. [2]. The analysis of the video content is a key step in the processing of videos. This step is carried out in two axes : video and audio [5]. Video analysis begins with a classification of video images (basic units). Then, a phase of calculating of some low level features is performed to extract the semantics from soccer video. In the video classification phase, two approaches are considered [8] : semantic and syntactic classifications. The semantic classification identifies two states of the game : the Play state and the Break state. The Play state occurs when the ball is in the field and the game is going on whereas the Break state is when the ball exceeds the goals or dab lines or if it is in the air or when the game is stopped. The syntactic classification is referenced by the production style of the game and the edition patterns of the video that helps the viewer to understand and appreciate the game. Audio analysis distinguishes two classes [9] : action and non action. The action class contains highlights while the non action class includes silence, comments and noise. Soccer video highlights are, thus, characterised, in the audio class, by the excitement of the commentator's voice with the noise of the crowd and in the visual class by a succession of long shots followed by medium shots and after by several close-up shots. The search for the dominant color in video key images is a key element in visual classification. The ratio of the dominant color is calculated by the following equation [2] :

$$\Pi = \frac{P_g}{P} \ . \qquad [1]$$

with P_g the number of green pixels and P the number of pixels in the image.

3. Highlights modelisation with HMMs

In an HMM, we have three states : observed or entry states (video states), hidden states (patterns states modellized) and output states. Hidden states are obtained by a learning process effected on a base which contains many video sequences issued from several soccer matches. System states are inter-related by connections. A connection between a hidden state and an observed state represents the generation probability of an observed state taking into consideration that the Markov process is in a particular hidden state. The Markov model is also represented by a matrix of probabilities called "probability matrix of symbols" or "confusion matrix". The matrix contains the occurrence probabilities of the observed state with a particular hidden state [2].

In what follows, we will present the elements of our HMM. The states of our model are types of shots of images extracted from soccer video :

- State L for a long shot ;
- State M for a medium shot ;
- State C for a close-up shot.

An HMM is composed of an observation set. This set contains a sequence of shots of various highlights which constitutes our learning base. The HMM also comprises a matrix of observations B whose values are calculated by the equation given below :

$$B = \begin{pmatrix} P(o_t = L/q_t = L) & P(o_t = L/q_t = M) & P(o_t = L/q_t = C) \\ P(o_t = M/q_t = L) & P(o_t = M/q_t = M) & P(o_t = M/q_t = C) \\ P(o_t = C/q_t = L) & P(o_t = C/q_t = M) & P(o_t = C/q_t = C) \end{pmatrix} \quad [2]$$

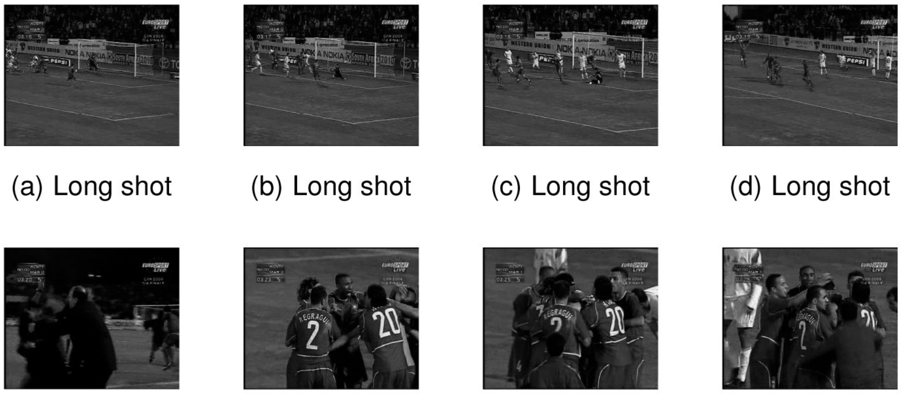
$P(o_t = L/q_t = M)$ is the probability of observing state L taking into account that we are in the medium shot. The third element of our HMM is the matrix of transitions A. Their values are calculated by the equation below :

$$A = \begin{pmatrix} P(q_{t+1} = L/q_t = L) & P(q_{t+1} = L/q_t = M) & P(q_{t+1} = L/q_t = C) \\ P(q_{t+1} = M/q_t = L) & P(q_{t+1} = M/q_t = M) & P(q_{t+1} = M/q_t = C) \\ P(q_{t+1} = C/q_t = L) & P(q_{t+1} = C/q_t = M) & P(q_{t+1} = C/q_t = C) \end{pmatrix} \quad [3]$$

with $P(q_{t+1} = M/q_t = L)$ the probability to pass from state L to state M. The fourth element of our HMM is the set of initial probabilities. The probabilities are calculated by the following equation :

$$\Pi = \begin{pmatrix} P(q_1 = L) \\ P(q_1 = M) \\ P(q_1 = C) \end{pmatrix} \quad [4]$$

with $P(q_1 = L)$ the probability to have the long shot L in the first state of HMM. The last elements of our model are exit states : highlight and common event. Figure 1 presents a sequence of shots of a soccer video highlight.



(a) Long shot (b) Long shot (c) Long shot (d) Long shot

(e) Close-up shot (f) Close-up shot (g) Close-up shot (h) Close-up shot

Figure 1. *A sequence of shots of a soccer video highlight*

4. Learning and recognition algorithms in HMMs

The recognition of highlights in video soccer by using HMMs is realized in two steps. A first step for the learning of the HMM and a second one for the recognition of highlights. In the learning of HMM phase, we have used a database of video sequences of highlights. The learning give final values to matrixes A and B and to the vector Π of HMM. The resultant HMM constitutes the main element of the recognition phase because the HMM take in entry vectors of observable symbols and by using estimation algorithms it calculates the probability that input sequences belong to the highlights class.

The processes of learning and recognition of highlights in video soccer are carried out by using Estimation of Distribution Algorithms (EDA) : Forward-Backward and Viterbi. They are inspired from genetic algorithms and they are used to solve optimization problems by the applying of a sampling of the function which describe the quality of possible solutions. The Forward-Backward algorithm calculates the likelihood probability of a sequence of observations in a given HMM. The algorithm is able to determine the most likely model to generate a sequence of observations in terms of maximum likelihood. Among all the possible paths, the Viterbi algorithm delivers, the path that corresponds to the most probable sequence of shots in terms of the likelihood probability of a sequence of observations O. It also calculates the likelihood probability on the best path. Taking into consideration that we have a Hidden Markov Model, the calculation of the probability of having an observation is done by the Forward-Backward algorithm [6, 7]. After the operation of extracting the possible paths of states, the choice of the best one is a classic problem linked to Hidden Markov Models and it's solved by the Viterbi algorithm. Given an observation and an HMM, the Viterbi algorithm determines the best path of states [6].

5. Experimental results

In our study, we have extracted highlights from different soccer videos with different probability values. We applied on them the learning process to obtain the values of the matrixes A and B and the vector Π of our HMM. The resultant HMM calculates the probability that an observation sequence belong to the highlights class. We have used two databases to envelope our system. A first highlights database which constitutes the learning database of our HMM. The other database is used in recognition of highlights and non highlights in soccer videos. Video sequences of both, learning and recognition databases, are issued from different soccer videos.

We have applied our model on events in MPEG format 352*288 with a throughput of 1150 kbps. The learning process with the HMMs is undertaken on a base of 70 highlights. The size of the events is variable and it depends on the importance given to the event by the video producer. The recognition process is carried out on 271 video segments of various soccer videos to demonstrate the effectiveness of our system. The tests' base contains 165 highlights and 106 common events. The observations set V is composed of a set of subsets whose size ranges from 4 to 10 states. The shots' sequence of the previous highlight corresponds to the following observations set : $[LLLLCCCCC]$ with L : the long shot and C : the close-up shot. The probability of plausibility on the best path is equal to 0.0160. This value is above the threshold 0.001. The classification phase of images is carried out by Y. Tabii et al. [1]. They proceed to the segmentation of images by the division of shots in format 3 :5 :3. Then, they classify images by their types of shots. The experimental results of this new method are very promising and improve the performance of the shots classification.

We have tested our Hidden Markov Model on a database of 271 sequences issues from video unlike from those used during the learning phase. Table 1 illustrates the results of the classification with HMMs. It gives the number of video sequences well classified and the number of sequences badly classified.

Tableau 1. *Results of observations' classification by the HMMs*

Sequences	Good classification	Bad classification	Total
Highlights	a=156	b=9	165
Common events	c=94	d=12	106
Total	250	21	N=271

The accuracy is calculated by the following formula :

$$P = a/a + b = 156/165 = 0.945 \quad [5]$$

The value of 0.945 shows that the model issued from the learning phase gives a good ratio of succeed classification.

Table 2 shows the classification results of different type of highlights : goal, corner kick, direct free kicks,...It also presents the percentages of correct events, missed and wrong ones. The use of HMMs gives a high value of the percentage of correct classification.

Tableau 2. *Results of the use of HMMs classified by highlights categories*

Sequences	Detected	Correct	Missed	False
Goal	66	60(90,9%)	1(2,2%)	6(9%)
Corner kick	9	9(100%)	0(0%)	0(0%)
Direct free kicks	63	60(95,2%)	1(2,3%)	3(4,7%)
Yellow/red card	12	12(100%)	0(0%)	0(0%)
Penalty	15	15(100%)	0(0%)	0(0%)

6. Related works comparison

We have compared, in table 3, our results with others by using other methods concerned with the extraction of video soccer highlights.

Tableau 3. *Comparative table of results achieved by the use of BN/DBN, the appearance of the referee and HMMs*

Sequences	Goal	Corner kick	Penalty	Yellow/red card	Direct free kick
Accuracy (Bertini, Finate state machine models)	—	94,7%	94,7%	—	94,7%
Accuracy (Ekin, lines detection)	—	—	100%	100%	62%
Accuracy (HMM)	90,9%	100%	100%	100%	95,23%

— means that the author of the classification method haven't gave results related to the corresponding highlight category.

The comparison of our results with those presented by the authors of the two proposed methods shows that the highlights recognition with Hidden markov Models give good results. Our recognition database is different from those used by M. Bertini et al.'s [4], who used engine models with finite states to model highlights and by A. Ekin et al. [2] who employed the detection of goal lines and the detection of a game break, in the sequences classification into highlights or non highlights classes.

7. Conclusion

In this paper, we have presented tools for the segmentation and classification of soccer video. We have used long, medium and close-up shots as basic semantic elements. We have also used domain knowledge to get highlights from soccer video. Then, we have carried out the classification and segmentation by using HMMs followed by dynamic programming. Hidden Markov Models offer the advantage of having a short computational time and a high successful detection rate. The obtained results show that high level video structures give a good accuracy during the classification. The main contribution of this paper is the use of temporal properties of video scenes to improve the accuracy of interpretation of soccer video semantics.

8. Bibliography

[1] Y. TABII, M. OULD DJIBRIL, Y. HADI, R. OULAD HAJ THAMI, « A New Method for Video Soccer Shot Classification », *CORESA07, VISAPP (1)*, vol. 1, pp. 221–224, 2007.

[2] A. EKIN, A. M. TEKALP, R. MEHROTRA, « Automatic Soccer Video Analysis and Summarization », *IEEE Transactions on Image Processing*, vol. 12(7), pp. 796-807, July, 2003.

[3] C.-L. HUANG, H.-C. SHIH, C.-Y. CHAO, « Semantic Analysis of Soccer Video Using Dynamic Bayesian Network », *IEEE Transactions on Multimedia*, vol. 8(4), pp. 749-760, August 2006.

[4] M. BERTINI, A. D. BIMBO, W. NUNZIATI, « Model Checking for Detection of Sport Highlights », *MIR03 ACM*, pp. 215-222, Berkeley, California, USA, November, 2003.

[5] E. KIJAK, « Structuration Multimodale des Vidéos de Sports par Modèles de Markov cachés », *book*, Institut de Recherche en Informatique et Systèmes Aléatoires, (IRISA), Décembre, 2003.

[6] R. DUGAD, U. B. DESAI, « A Tutorial on Hidden Markov Models », *Technical report, SPPAN-96.1*, May, 1996.

[7] M. TEN BRUGGENCATE, S. CHALASANI, « Parallel implementations of the power system transient stability problem on clusters of workstations », *Supercomputing '95 : Proceedings of the 1995 ACM/IEEE conference on Supercomputing (CDROM)*, New York, NY, USA. p. 34, 1995.

[8] L. XIE, P. XU, S.-F. CHANG, A. DIVAKARAN, H. SUN, « Structure Analysis of Soccer Video with Domain Knowledge and Hidden Markov Models », *Pattern Recognition Lettres*, vol. 25, pp. 767-775, 2004.

[9] H. HARB, « Classification du Signal Sonore en Vue d une Indexation par le Contenu des Documents Multimédias », *book*, Lyon, 2004.

[10] M. ABED CHAIB, « Étude et Mise en oeuvre des Modèles de Markov Cachés en vue de la Reconnaissance de Trajet », Décembre, 2004.

[11] C. LU, M. S. DREW, J. AU, « An Automatic Video Classification System Based on a Combination of HMM and Video Summarization », *International Journal of Smart Engineering System Design*, Vancouver, B.C., Canada, 2002.

9. Biography

Rajae El Ouazzani : Master's in Information Technology and Communication (TIC) from the Faculty of Sciences and Technologies of Fes, Morocco in 2004. Diploma of Higher Education Approfondies (DESA) in Computer Sciences and Telecommunications from the Faculty of Sciences of Rabat, Morocco in 2006. PhD student in the 2nd year in the team WIM of SI2M Laboratory, Dept. SI3M Doctorate in the Ecole Nationale Superieure d'Informatique et d'Analyse Systems (ENSIAS) in Rabat, Morocco.

Rachid Oulad Haj Thami : Doctor in Computer Sciences from the University Hassan II in 2002. Authorized professor in the speciality : Databases and Multimedia databases. Team leader of WIM in SI2M Laboratory and Co Head of the Department Doctorate SI3M. Research field : Image Processing with segmentation, analysis and indexing information contained in the image.