

# Soccer Video Edition Using a Multimodal Annotation

Fendri Emna, Ben-Abdallah Hanène, and Ben-Hamadou Abdelmajid

**Abstract**—In this paper, we present an approach for soccer video edition using a multimodal annotation. We propose to associate with each video sequence of a soccer match a textual document to be used for further exploitation like search, browsing and abstract edition. The textual document contains video meta data, match meta data, and match data. This document, generated automatically while the video is analyzed, segmented and classified, can be enriched semi automatically according to the user type and/or a specialized recommendation system.

**Keywords**—XML, Multimodal Annotation, recommendation system.

## I. INTRODUCTION

OVER the last decade, there has been a rapid proliferation of multimedia applications to access and analyze multimedia documents. Several methods were proposed to extract efficiently low level information using MPEG'7 descriptors [1]. Other methods tried to extract semantic information from low level analyses in specific types of videos such as news [2], sports video [3], or films [4]. The majority of these methods offer limited solutions within a constrained context. For example, Pingali et al. [5], Sudhir [6] and Delakis [3] proposed systems for tennis video analyses; these systems rely on the nature of this game to segment the tennis video into sets, games and points. Peng and Snoek [7], Ekin [8], Xu et al. [9] proposed solutions to segment soccer video and extract some semantic features. Again, the proposed methods are specific to certain events/aspects of this game. For instance, Ekin [8] classifies the soccer video segments as long, medium, long or close-up shots in order to extract information used for player detection/tracking and referee detection.

In an effort to provide for a generic framework that can be reused to develop multiple applications, we have proposed FOCSE, a Framework for sOCcer Semantic Edition [10].

FOCSE includes the components/methods common to any semantic exploitation of soccer video: segmentation, segment classification and basic end-user applications.

Fendri Emna, Institut Supérieur des Etudes technologiques de Sfax, Tunisie (e-mail : Mir@cl laboratory ; fendri.msf@gnet.tn).

Ben-Abdallah Hanene, Faculté des Sciences Economiques et de Gestion de Sfax (e-mail: Mir@cl laboratory; hanene.benabdallah@fsegs.rnu.tn).

Ben-Hamadou Abdelmajid, Institut Supérieur d'Informatique et multimedia, Sfax, Tunisie (e-mail : Mir@cl laboratory, abdelmajid.benhamadou@gmail.com).

The segmentation adopts a generic method that relies on graphic low level features (luminance variability) [11]. On the other hand, segment classification maps additional low level features (e.g. color, shape, etc.) into semantic information specific to soccer, e.g., region production effects, person's detection, events, etc.

Finally, FOCSE offers the three basic end-user applications: indexing, search and summarization.

These basic applications can be included in other complex applications such as web search engine, interactive TV, etc.

Being a framework, FOCSE, proposes both an architectural design for the common methods, and library code that can be reused to derive other applications for a semantic exploitation of soccer videos.

Similar to currently proposed end-user applications, the underlying methods of FOCSE are based on graphic low-level features. However, such an approach is efficient for a limited number of semantic applications. In fact, several semantic features either have no representing low-level descriptors (e.g. player's emotion) or they are too complex to be represented solely by graphic descriptors (e.g. ball exchange, a particular player). This limit of graphic descriptors motivated us to extend the segmentation and the classification method to profit from a second modality: the text. Such an extension will, on one hand, increase the expressive power of segmentation and classification; on the other hand, it will increase the efficiency of end-user applications, e.g., a better response-time, recall and precision rates of search engines.

In this paper, we propose to annotate soccer matches through a structured textual document in analogy with scripts for films. The annotation document can be used for search and browsing ends, etc. It contains information about the video file, game meta data and the game itself. It is semi-automatically generated as an XML file: the automatically detected features are the basic structure that can be further enriched manually with the help of a specialized recommendation system.

The remainder of this paper is divided into three sections. In section II, we present the DTD for soccer video documentation. Section III describes the steps for soccer document generation and enrichment. Section IV summarizes the presented work and outlines future work.

## II. THE SOCCER DOCUMENT STRUCTURE

A DTD (Document Type Definition) is a typical structure

that an XML document must respect to be considered as valid [12] [13]. The DTD presented in Fig. 1 shows the typical

structure we propose to document soccer videos.

|  |   |
|--|---|
| <pre> &lt;?xml encoding="ISO-8859-1"?&gt;  &lt;!--ELEMENT DocumentMatchFoot (Metadatamatch,Metadatafile,Game)--&gt; &lt;!--ELEMENT Metadatamatch (Date, StartTime, Place, Team1,Team2,score [2], Arbitration, Context)--&gt; &lt;!--ELEMENT Metadatafile(Path, FileName, Extension, Type, FrameNb, SegmentNb)--&gt; &lt;!--ELEMENT Game (Segment*)--&gt; &lt;!--ELEMENT Segment (SegmentBegin, SegmentEnd, Class, Player*, Event*)--&gt; &lt;!--ELEMENT Date (EMPTY)--&gt; &lt;!--ATTLIST Date Day CDATA #REQUIRED Month CDATA #REQUIRED Year CDATA #REQUIRED--&gt;  &lt;!--ELEMENT StartTime (EMPTY)--&gt; &lt;!--ATTLIST StartTime Hour CDATA #REQUIRED Minute CDATA #REQUIRED --&gt;  &lt;!--ELEMENT Place (EMPTY)--&gt; &lt;!--ATTLIST Place Field CDATA #required Town CDATA #required Country CDATA #IMPLIED--&gt; &lt;!--ELEMENT Team1 (Team)--&gt; &lt;!--ELEMENT Team2 (Team)--&gt; &lt;!--ELEMENT Team (TeamName, ShortColor, TeeshirtColor, Coach, PlayerIn, PlayerOut, PlayerToIN)--&gt;  &lt;!--ELEMENT TeamName (EMPTY)--&gt; &lt;!--ATTLIST TeamName Name CDATA #REQUIRED Abbreviation CDATA #IMPLIED--&gt;  &lt;!--ELEMENT ShortColor #PCDATA--&gt; &lt;!--ELEMENT TeeshirtColor #PCDATA--&gt;  &lt;!--ELEMENT Coach (EMPTY)--&gt; &lt;!--ATTLIST Coach CoachFirstName CDATA #REQUIRED CoachNameCDATA #REQUIRED nationality CDATA #IMPLIED--&gt; &lt;!--ELEMENT PlayerIn (Player)+--&gt; &lt;!--ELEMENT PlayerOut (Player)*--&gt; &lt;!--ELEMENT PlayerToIn (Player)+--&gt; &lt;!--ELEMENT Player (EMPTY)--&gt; &lt;!--ATTLIST Player FirstName CDATA #REQUIRED Name CDATA #REQUIRED Number CDATA #REQUIRED YellowCard CDATA #IMPLIED Nationality CDATA #IMPLIED--&gt; &lt;!--ELEMENT Score #PCDATA--&gt; &lt;!--ELEMENT Arbitration (Referee1,Referee2, Referee3, Referee4)--&gt; &lt;!--ELEMENT Referee1 (Referee)--&gt; &lt;!--ELEMENT Referee2 (Referee)--&gt; &lt;!--ELEMENT Referee3 (Referee)--&gt; &lt;!--ELEMENT Referee4 (Referee)--&gt; &lt;!--ELEMENT Referee (EMPTY)--&gt; </pre> | <pre> &lt;!--ATTLIST Referee FirstName CDATA #REQUIRED Name CDATA #REQUIRED Nationality CDATA #IMPLIED--&gt;  &lt;!--ELEMENT Context #PCDATA--&gt; &lt;!--ELEMENT Path (#PCDATA)--&gt; &lt;!--ELEMENT FileName (#PCDATA)--&gt; &lt;!--ELEMENT Extension (#PCDATA)--&gt; &lt;!--ELEMENT Type (#PCDATA)--&gt; &lt;!--ELEMENT FrameNb (#PCDATA)--&gt; &lt;!--ELEMENT SegmentNb (#PCDATA)--&gt; &lt;!--ELEMENT SegmentBegin (#PCDATA)--&gt; &lt;!--ELEMENT SegmentEnd (#PCDATA)--&gt;  &lt;!--ELEMENT Class (EMPTY)--&gt; &lt;!--ATTLIST Class Fieldclass (Field   Gallery  Center PenaltyBox   advertisingPanel  TeamSeat) ScoreClass (Score  NoScore) AnnouncementClass (None  Error  Warning  Replacement  Statistics)--&gt;  &lt;!--ELEMENT Event (Mistake?, Pass*,corner?, Error?, touch?, Goal?, Appears*, Penalty?, Exchange?, FreeKick?, Replacement?, Offside?, Tacle ?, Attack?, Other*)--&gt; &lt;!--ELEMENT Mistake (Player*, MistakeType)--&gt; &lt;!--ELEMENT MistakeType (empty)--&gt; &lt;!--ATTLIST MistakeType (Simple  Hand  SeriousSanctionned   SeriousNoSanctionned)--&gt; &lt;!--ELEMENT Passe (From,To)--&gt; &lt;!--ELEMENT From (Player)--&gt; &lt;!--ELEMENT To (Player)--&gt; &lt;!--ELEMENT Corner (Position, Player)--&gt; &lt;!--ELEMENT Position (Left Right)--&gt; &lt;!--ELEMENT Error (ErrorType,Player)--&gt; &lt;!--ELEMENT ErrorType (Error Warning)--&gt; &lt;!--ELEMENT Touch (Player)--&gt; &lt;!--ELEMENT Goal (Player)--&gt; &lt;!--ELEMENT Appears (Player)--&gt; &lt;!--ELEMENT Penalty (Player)--&gt; &lt;!--ELEMENT Exchange (J1,J2,PassNb)--&gt; &lt;!--ELEMENT Tacle (J1,J2)--&gt; &lt;!--ELEMENT J1 (player)--&gt; &lt;!--ELEMENT J2 (player)--&gt; &lt;!--ELEMENT Attack (EMPTY)--&gt; &lt;!--ATTLIST Attack from CDATA #IMPLIED TO CDATA #IMPLIED--&gt; &lt;!--ELEMENT PassNb (#PCDATA)--&gt; &lt;!--ELEMENT FreeKick (FreeKickType , Player)--&gt; &lt;!--ELEMENT FreeKickType (EMPTY)--&gt; &lt;!--ATTLIST FreeKickType (Direct Nodirect)--&gt; &lt;!--ELEMENT Relpcement(TeamName, Out, In)--&gt; &lt;!--ELEMENT Out (player)--&gt; &lt;!--ELEMENT In (player)--&gt; &lt;!--ELEMENT offside (EMPTY)--&gt; &lt;!--ATTLIST offside (Reported NoReported)--&gt; &lt;!--ELEMENT Autre (#PCDATA)--&gt; </pre> |
|--|---|

Fig. 1 DTD for the XML soccer document

In this structure, the soccer video document is composed of three parts. The first one describes the video file's meta data; the second one describes the game's meta data, and the last one contains information about the game progress.

#### A. Meta Data of the Video File

These meta data gathers data about the soccer video file; in particular, its name, path, type, duration and eventually the number of detected video segments.

#### B. Meta Data of the Game

We mean by game meta data all data about the match: date and hour of beginning of the match, the teams, the referees, the context of the game (e.g., national cup, national championship, international manifestation, etc), and eventually the score which can be, by itself, sufficient for some statistical end applications.

#### C. Game Progress Data

In this part of the document, we describe information about the content of the game, judged useful for further semantic explorations. For each segment of the video, we can precise its class, the present players, and particular events detected. We single out as particular events mistakes, a pass, corners, errors, touch, goal, penalty, exchange, free kick, replacement, offside, or other event to be specified through a manual annotation.

Furthermore, most of the document elements can be automatically generated through (graphic) low-level analyses. The remaining elements (marked in bold face in Fig. 1) require manual intervention.

Note that our documentation structure can be reused to document other types of games. In fact, it includes a large portion that is generic; only the elements starting from "Event" are specific to soccer.

### III. SOCCER DOCUMENT GENERATION

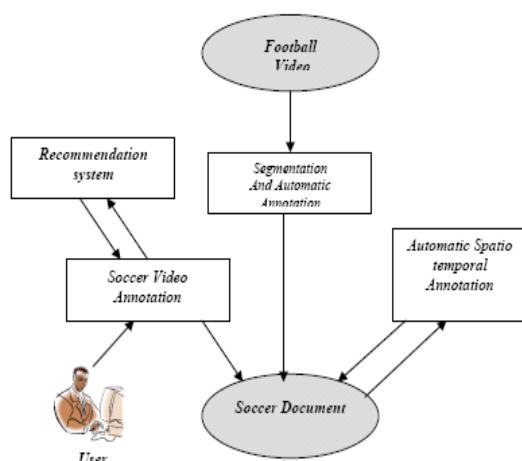


Fig. 2 Soccer Document Generation

As illustrated in Fig. 2, the generation of a soccer document is done in three steps. The first step is entirely automatic using

personalized MPEG'7 descriptors while segmenting the video stream. It generates the main structure of the game progress part of the document. The second step semi automatically annotates the main structure with additional features. This step depends on the user type and relies on an appropriate recommendation system. The last step is entirely automatic and enriches the soccer document by new information. This enrichment uses contextual soccer rules.

#### A. Automatic Soccer Video Annotation

The first step in a soccer document generation is done in parallel with a graphic-level video segmentation. In this low-level segmentation, we divide the soccer video into different segments using a specific shot detection algorithm [11]. In addition, we also have developed personalized MPEG'7 descriptors for some semantic feature extraction.

This first step produces the content of the tags marked in bold in Fig. 1: they correspond to file meta data, segmentation information, classification of segments, a part of segment classification and several events like replacement, attack and errors' announcements. Fig. 3 partially presents an example of an XML soccer document created after this step of video low level video analyses.

```

- <Game>
- <metadatafile>
  <path>'c:\EuropeCup'</path>
  <FileName>'EspagneVsAllemagne'</FileName>
  <Extension>'avi'</Extension>
  <Type>'video'</Type>
  <FramesNbre>'135000'</FramesNbre>
  <SegmentNbre>'1245'</SegmentNbre>
</metadatafile>
- <Segment>
  <Segmentbeginning>20</Segmentbeginning>
  <SegmentEnd>150</SegmentEnd>
  <Class fieldClass="PenaltyBox" ScoreClass="Noscored"
    AnnoucementClass="none" />
- <Event>
  <attack from="50" TO="100" />
</Event>
</Segment>
...
- <Segment>
  <Segmentbeginning>2200</Segmentbeginning>
  <SegmentEnd>2350</SegmentEnd>
  <Class fieldClass="Gallery" ScoreClass="scored"
    AnnoucementClass="none" />
</Segment>
...
</Game>
  
```

Fig. 3 An example of an automatically generated soccer document

## B. Semi Automatic Annotation of the Soccer Video

Once the first version of the soccer video document is generated, users can add further annotations pertinent to future exploitation. This annotation depends on the user type and it is supervised by a recommendation system.

### 1) User Types

We can distinguish three types of actors in this step: video producers, match analyst and other users. These users do not have the same interests in a soccer video sequence.

While the game analyst can only update the match meta data and the match data, the simple user is simply authorized to update match data. On the other hand, the match producer is allowed to annotate different data. Fig. 4 represents the main tasks for the three types of users during the annotation step.

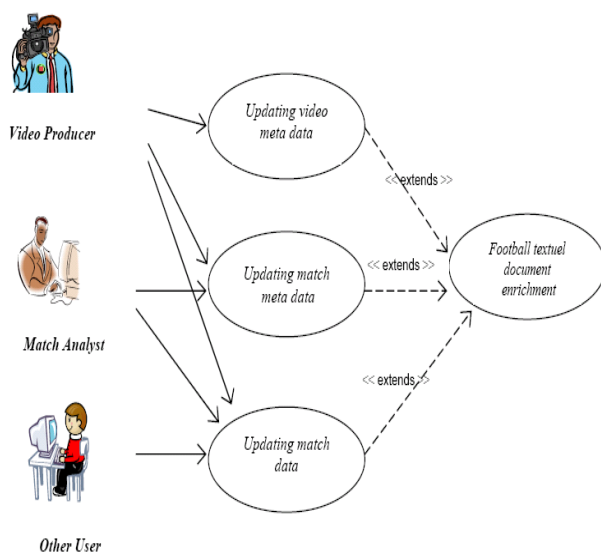


Fig. 4 Video soccer semi-automatic annotation

### 2) The Recommendation System for Soccer Annotation

Since the annotation is done manually, it seems beneficial to assist the annotator by providing him/her with recommendations to correctly annotate the video. For this, we propose to include a recommendation system for soccer video annotation.

This system offers a list of contextual recommendations. We next present a sample of recommendations depending on the user type:

- For all users, contextual annotations of the game progress are allowed by the recommendation system :
  - When annotating a segment whose region class is either "field" or "penalty box", the recommendation system proposes the insertion of any "event" tags and players from the two teams "PlayersIN" lists.
  - When annotating a segment whose region class is replacement player, the recommendation system proposes annotation with either "Players" from the two teams lists "PlayersToIN", or "Coaches".
  - When annotating a segment with a replacement event, the recommendation system first offers the list of "PlayersIN" to choose the player quitting the

match. Then, it allows the annotator to choose the new player from a restricted list composed of the "PlayersToIN" list of the replaced player's team.

- For video producers
  - the recommendation system offers them a maximum of annotation depending on the segment class.
- For match analysts
  - When the annotator is the match analyst, the recommendation system proposes annotations of match meta data

### 3) Automatic Enrichment of the Soccer Document

Several annotations can be added automatically after the second generation step. For example, suppose that an annotator has marked a segment with successive "pass" tags between two players (P1, P2) as shown in Fig. 5 where each arrow represents the ball itinerary. This scenario will be represented in the XML document as shown in Fig. 6.

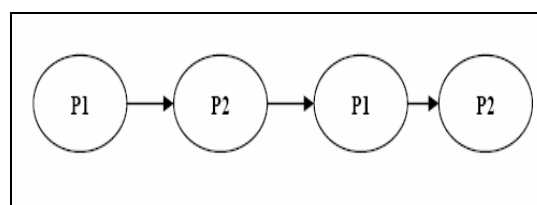


Fig. 5 An example of a successive pass scenario

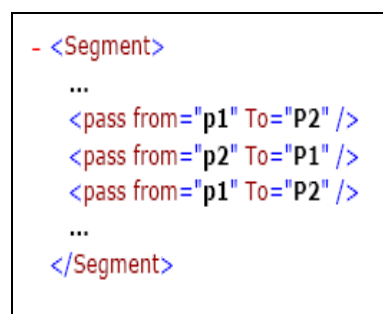


Fig. 6 Example of passes annotation

This annotation scenario can be used to automatically deduce two additional annotations:

- the presence of the players P1 and P2 in the segment;
- the presence of an exchange.

Note that the first derived annotation is difficult to realize with low level analyses. On the other hand, asking the annotator to manually annotate each segment with the present players may be a tedious task. Thus, the automatic enrichment of the document with this information is vital.

In this case, the automatic enrichment step will generate the enrich version of the XML document shown in Fig. 7.



```
- <Segment>
...
<pass from="p1" To="P2" />
<pass from="p2" To="P1" />
<pass from="p1" To="P2" />
<appears player="P1" />
<appears player="P2" />
<exchange J1="P1" J2="P2" PassNbre="3" />
...
</Segment>
```

Fig. 7 The resulting enriched document

Another example of an automatic enrichment of the soccer document can be the insertion of error (red card) for a player once he receives a second warning (two yellow cards). Fig. 8 shows the automatically enriched document for this scenario. Notice again the automatically derived annotation about the present players.

```
- <Game>
...
- <Segment>
...
<error errorType="warning" player="zidane" />
...
</Segment>
- <Segment>
...
<error errorType="warning" player="zidane" />
...
</Segment>
...
</Game>
```

Fig. 8 (a) Soccer document with user annotation

```
- <Game>
...
- <Segment>
...
<error errorType="warning" player="zidane" />
<appears player="zidane" />
...
</Segment>
- <Segment>
...
<error errorType="warning" player="zidane" />
<appears player="zidane" />
<error errorType="Error" player="zidane" />
...
</Segment>
...
</Game>
```

Fig. 8 b) Soccer document after automatic enrichment with error tag

#### IV. CONCLUSION AND FUTURE WORK

In this paper, we have first presented a structure for soccer video documentation; this structure contains three parts that interest three typical main types of users: 1) file meta data to mainly interest to video producers, 2) match meta data of interest to video producers and match analysts, and 3) game progress data of interest of all users. Secondly, we presented an approach for soccer document generation. This approach is composed of three steps. An automatic step that relies on low level information extraction from the game video; a semi automatic step for soccer document enrichment supervised by a recommendation system and depending on the user type; and finally, an automatic enrichment step that uses contextual rules to enrich the soccer document.

We are currently extending the recommendation system to provide for more coherence within the soccer document. We are also looking for other features for automatic enrichment of the soccer document. Finally, we are examining new low-level descriptors for a fine grain segmentation of the video through the detection of additional semantic features using a dynamic inference engine. These descriptors will be used both in the first step of document generation offering more significant information, and in the automatic document enrichment.

#### REFERENCES

- [1] <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>
- [2] Liu Huayong, Zhang Hui "The segmentation of news video into story units" Lecture notes in computer science ISSN 0302-9743 Congrès Advances in web-age information management : (6th international conference, WAIM 2005, Hangzhou, China, October 11-13, 2005 ) ISBN 3-540-29227-6.
- [3] E. Delakis « Structuration multimodale des vidéos de tennis en utilisant les modèles segmentaux », Thèse de doctorat, Université de Rennes 1, 2006.
- [4] Philippe Aigrain, Philippe Joly, "Automatic real-time analysis of film editing and transition effects and its applications", Computer & Graphics, 18(1), 1994, pp. 93-103
- [5] G. S. Pingali, Y. Jean, and I. Carlbom, "Real Time Tracking for Enhanced Tennis Broadcasts" Proc of IEEE CVPR (Computer Vision and Pattern Recognition), pp.260-265, 1998.
- [6] F. Sudhir, J. C. M. Lee and A. K. Jain, "Automatic classification of tennis video for high-level content-based retrieval" in Proc. Int. Workshop on Content-based Access of Image and Video Databases (CAIVD'98), 1998, pp.81-90.
- [7] Peng Xu, Lexing Xie, Shih-Fu Chang, Ajay Divakaran, Anthony Vetro, Huifang Sun "Algorithms and System for System for Segmentation and Structure Analysis in Soccer Video" IEEE International Conference on Multimedia and Expo, Tokyo, Japan Aug 22-25, 2001.
- [8] Ahmet Ekin and A. Murat Tekalp "A Framework for tracking and analysis of soccer Video" Proc. VCIP 2002 (Visual Communications and Image Processing). San Jose CA.
- [9] Algorithms and Systems for Segmentation and Structure Analysis in Soccer Video, IEEE International Conference on Multimedia and Expo, Tokyo, Japan, Aug. 22-25, 2001 P. Xu, L. Xie, S.-F. Chang, A. Divakaran, A. Vetro, H. Sun.
- [10] Fendri E., Ben-Abdallah H., Ben Hamadou A. "FOCSE : A framework for Soccer Video Edition" MediaNet 2004, Tozeur, Tunisie Dec. 2004.
- [11] Fendri E, Warda S., Ben Abdallah H. "Nouvelles méthodes de détection de transitions progressives dans la vidéo" SETIT, Sciences of Electronics, Technologies of Information and Telecommunications Sousse Tunisie, Mars 2003.
- [12] <http://www.info-3000.com/html/xml/dtd.php>
- [13] XML Schema Part 1: Structures Second Edition W3C Recommendation 28 October 2004, <http://www.w3.org/TR/2004/REC-xmlschema-1-20041028/>