2014 IEEE/ACM International Symposium on Big Data Computing

# Soccer Video Summarization using Video Content Analysis and Social Media Streams

Said Jai-Andaloussi
Faculty of sciences, Casablanca Hassan II University
LIAD Labs and GreenTic
Casablanca, Kingdom of morocco
said.jaiandaloussi@etude.univcasa.ma

Aboulaite Mohamed
Faculty of sciences, Casablanca Hassan II University
Casablanca, Kingdom of morocco

Nabil Madrane
Faculty of sciences, Casablanca Hassan II University
LIAD Labs and GreenTic
Casablanca, Kingdom of morocco
nabil.madrane@etude.univcasa.ma

Abderrahim Sekkaki
Faculty of sciences, Casablanca Hassan II University
LIAD Labs and GreenTic
Casablanca, Kingdom of morocco
abderrahim.sekkaki@etude.univcasa.ma

***Abstract*—Many televiewers and soccer fans choose to watch summaries of football games as watching a full soccer match requires a lot of time. Generally, the number of soccer matches broadcasted every weekend and during the week is very important. Taking the example of the most famous soccer leagues (Spain, England, Italy, Germany, French), we've ended up with fifty of soccer matches released every week (75 hours of soccer broadcasting): hence the need for developing a system of video summarization. We note that soccer video summarization is complete in the sense that it contains important events (significant action) extracted from soccer matches. Habitually, the soccer match summarization is done manually, however, this requires a considerable amount of time. For that reason, it is necessary to have a means for doing the soccer match summarization automatically. Several works have been developed for video soccer summarization. However, the previous approaches use only the video content to create a summary of the soccer video. Owing to the wide semantic disparity between low-level features and high-level events, it is not evident to come up with a generic model to achieve a high accuracy of video soccer summarization. In this paper, we present a novel approach for soccer video summarization based on video content analysis and social media streams. Social media streams such as Twitter generate important volume of content for most sport events on a daily basis. The mining of the tweets can be used for the detection of events, which can be qualified as soccer match highlights. Incorporating social media events detection into video content analysis significantly improves the quality of the soccer video summarization. Results of experiment applications are performed over a database, consisting of more than 30 hours of soccer video.**

## I. Introduction

With the rise of soccer content broadcasting, soccer fans often find themselves unable to watch live soccer matches for several reasons such as time differences between countries and the simulcast of several matches. Generally, only small bits of a soccer match are interesting enough and worth the highlight. Automatic soccer video summarization looks like a very important tool for soccer teams as well. Events during play can be highlighted and the video summary can be shown to the team or individual at half-time or after the game [1] to correct some game mistakes or to change the tactics of the game. In relation to this, soccer video summarization has lately attracted much research and a large spectrum of possible applications have been considered. Several approaches have been proposed in this field of research. Baoxin et al [2] proposed both deterministic and probabilistic approaches for the detection of the plays. The detected plays are concatenated to create summary of the original sports video. These approaches use high-level inference to make the final decision based on the low-level features (e.g. field colors and their spatial patterns, scene cuts, etc..). Ekin et al [3] introduced a fully automatic and computationally efficient framework for the summarization of soccer videos using cinematic and object-based features. The proposed method includes low-level soccer video processing algorithms, such as robust shot boundary detection and dominant color region detection, as well as some higher-level algorithms for goal detection and penalty-box detection. In other works, texture features and object color are used to create highlights [4] and to parse video soccer games [5]. Interactions and the trajectories of object motion are used for soccer play classification [6] and event detection [7]. Changsheng Xu et al [8] proposed an framework of live sports event detection based on incorporation of web-casting text into sports video analysis. Zhao [9] et al introduced a highlight summarization of sports video based on audio energy and motion activity. Nichols et al. [10] detect and summary the highlight in soccer video using only Twitter data.

In this paper, we propose a new framework for automatic soccer video analysis and summarization by using video content analysis and social media streams (e.g. Twitter). An organization chart of the proposed framework is given in

978-1-4799-1897-3/14 $31.00 © 2014 IEEE
DOI 10.1109/BDC.2014.20

Figure 1.

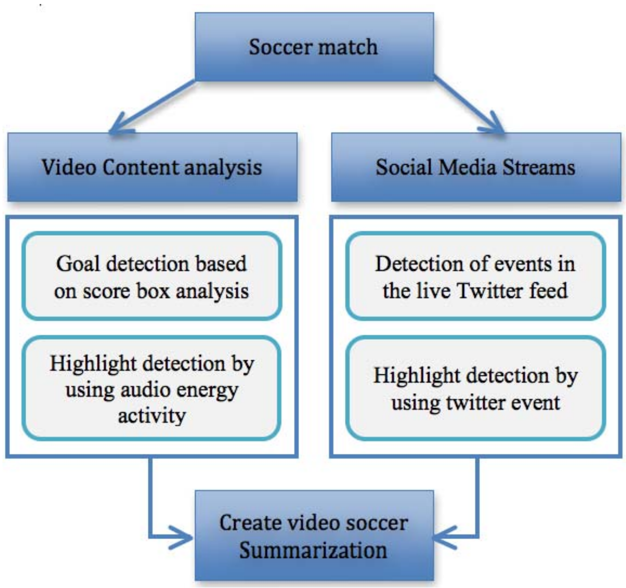


Figure 1. Organization chart of Soccer Video Summarization

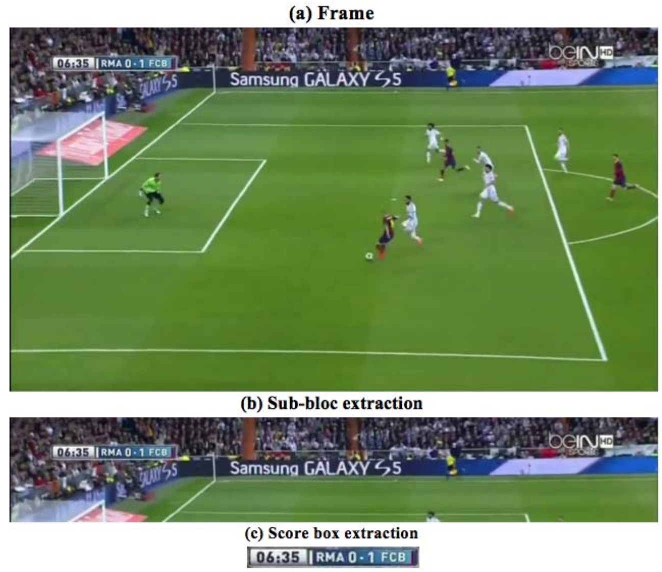


Figure 2. Flowchart of score-box localization and goal detection

## II. VIDEO CONTENT ANALYSIS

In this section, we will detail the part of the content analysis of soccer video that allows the detection of highlights in the soccer games by using two different methods: 1) the first based on the study and the analysis of score-box in soccer video, in order to detect the goals in a soccer games; 2) the second based on the analysis of the audio energy activity in the commentators speech within the objective of detection of highlights in the soccer games

### A. *score-box detection and extraction*

In broadcast soccer videos, a superimposed score-box is used to display game status (e.g. score, team names, etc.), to rise the audiences understanding of the game progression. Additionally, the score-box changes after a goal event occurs. For that reason, localization and recognition of the score-box is very important for soccer video analysis, for example, as a method for detecting score events or as a source of evidence for a score detection. The principal stages of our algorithm for score-box localization and goal detection in soccer match video consist in extracting a clip from the full-length of the soccer match video, the length of the video clip extracted is $L$. Then we extract the score-box from sub-block (Fig 2 (b)) based on the following extracted proprieties: 1) texts remain stable in the same place during the match; 2) texts are superimposed in top/bottom of video frames. In the first propriety, we use motion vector to locate/extract texts and in the second, we select the sub-block (Fig 2) to investigate texts and to reduce the computation in motion vector algorithm. After that, we use the motion vector to detect unchangeable blocs/pixels on each frame by comparing them using the following condition:

$$\begin{cases} p(i,j)_{frame_{k-1}} = p(i,j)_{frame_k} \pm t \\ \Rightarrow \;\; S_L = S_L \cup \{p(i,j)_{frame_{k-1}}\} \end{cases} \quad (1)$$

where $p(i,j)$ is the pixel of the frame index $k$ in video clip $L$.. $S_L$ represents the set of pixels that form the score-box and $t$ denotes the threshold used to set the difference degree between two pixels of the same position in two successive frames. In our case $t$ is equal to 5.

By applying this condition on all frames extracted from the clip video $L$, the final result contains only the fixed pixels/block. By summing these pixels we localize the researched score-box as shown in the figure 2 (c).

### B. *Goal detection*

After score-box extraction, we perform a histogram equalization and a binarization (Thresholding) process by using the Otsu method [11] on the extracted score-box image $src$. Whose purpose is to normalize the brightness and increase the contrast of the $src$ image. After the preprocessing step, the final one to detect goals is by computing the total pixel change between two scores-boxes. For that, we have fixed two thresholds: $t_1$ and $t_2$. A goal is detected only if the total pixel change $t$ is between $t_1$ and $t_2$. If $t$ is greater than $t_2$ it is likely that the score-box has been moved due to compression methods or has been totally dropped (half time for example). When a goal is detected we initialize the score-box with the new one. We repeat this process until the end of the input video. The following flow chart shows the proposed algorithm:

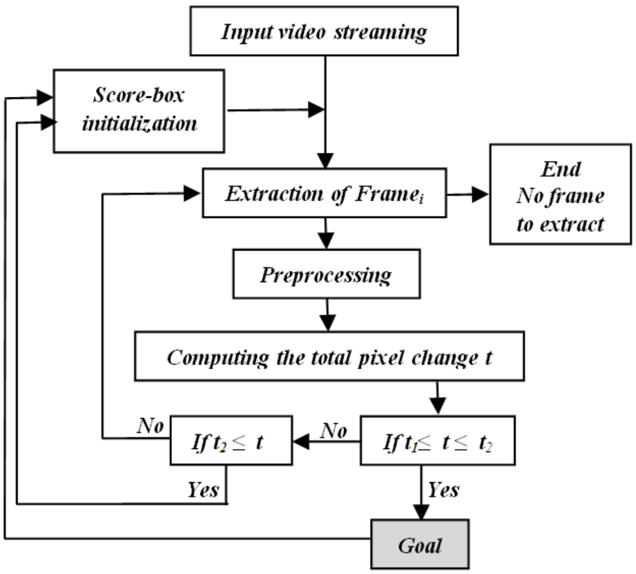


Figure 3. Chart of the algorithm used for score-box localization and goal detection

*C. Highlight extraction using audio features*

Adding to the visual information in video track, the Audio track also wears important information that should not be overlooked and, also, gives the semantic to the video, particularly in action and sport videos. In the case of soccer game, when there is an important moment (goal, penalty, etc..), the voice intensity of the commentators rises proportionally with the action. The information present in the intensity of the audio track can be used as descriptor To further characterize the video.

In this paper, we proposed a new approach for highlights extraction using audio features. The first step in our new method for highlight extraction is to isolate audio stream from the video stream. Then we perform an EMD decomposition [12] to the audio signal to break down a signal without leaving the time domain. The EMD is locally adaptive and suitable for analysis of non-linear or non-stationary processes. The starting point of EMD is to consider oscillatory signals at the level of their local oscillations and to formalize the idea that: Signal = fast oscillations surimposed to slow oscillations and to iterate on the slow oscillation components considered as a new signal. This one-dimensional decomposition technique extracts a nite number of oscillatory components or well-behaved $AM - FM$ functions, called intrinsic mode function ($IMF$), directly from the data. Given a signal $x(t)$, the effective algorithm of EMD can be summarized as:

1) Identify all extrema of $x(t)$
2) Interpolate between minima (resp. maxima), ending up with some envelope $emin(t)$ (resp. $emax(t)$)
3) Compute the mean $m(t) = (emin(t) + emax(t))/2$
4) Extract the detail $d(t) = x(t) - m(t)$
5) Iterate on the residual $m(t)$

In practice, the above procedure has to be refined by a sifting process [12] which amounts to first iterating steps 1 to 4 upon the detail signal d(t), until this latter can be considered as zero-mean according to some stopping criterion. Once this is achieved, the detail is referred to as an Intrinsic Mode Function (IMF), the corresponding residual is computed and step 5 applies. By construction, the number of extremas decreased when going from one residual to the next, and the whole decomposition is guaranteed to be completed with a finite number of modes.

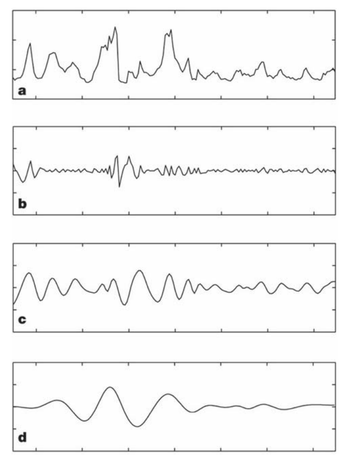


Figure 4. (a) Original signal (b) first IMF (c) second IMF (d) third IMF

After the extraction of IMFs, the next step is to choose the most faithful IMF representing the input signal, then based on Persistent 1D algorithm [12] we extracted the local maxima of the selected IMF. Those maxima consist of the soccer video highlights as shown in figure 5.

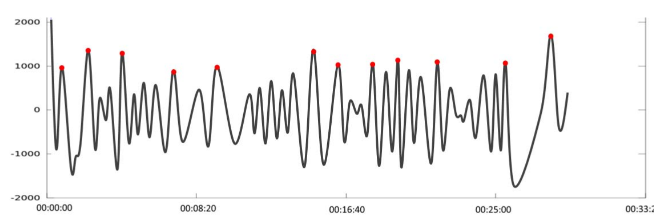


Figure 5. Detection of IMF maxima that represents the highlights of video soccer

The final step is to extract all maxima from the original video to build the highlight summarization. But the only problem that arises is the commentators speech that changes from commentator to another depending on their respective languages and nationalities.. There are, in general, 2 types of commentators schools, the Latina/Arabian school and the Europeans school. Thus, the style chosen by the commentator widely affects the efficiency of our audio-based approach for highlights extraction. For this reason, we will add other types of information that can confirm the result obtained using the audio analysis and also improve the quality of the highlights extraction. That leads us to use social networking and microblogging services such as Twitter.

## III. SOCCER HIGHLIGHTS DETECTION BASED ON ANALYZING TWEETS OF TWITTER

In this section, we present our proposed approach for soccer game summarization using tweets of Twitter. Every hour of the day, internet users are posting millions of status updates to social networks, such as Twitter and Facebook. As of January 2014, more than 9,100 updates were being posted to twitter each second (e.g. in the 2014 world cup a new record has been reached: 35.6 million tweets were sent during the 90-minute game). Analysis of the amount of posted tweets and, especially, the number of tweets versus time give an important information about what is happening in real-time.

In this work, we use two methods to detect the soccer game highlights. In the first method, we propose moving threshold burst detection to detect bursts of tweets on Twitter [14]. These bursts can be considered as the highlights in the soccer game. For that, we use moving-threshold to detect the points of highlights happening in the soccer games. After the highlights detection, in the second step, we extract the semantics of the detected highlights from the tweets.

### *A. Burst detection based on moving-threshold*

Many previous methods of burst detection use predefined threshold value before detecting possible bursts. The choice of the threshold significantly affects the result of burst detection. Therefore, the choice of the threshold becomes an important step since for the real-time event detection of soccer game, it is very difficult to determine a threshold previously.

In this paper, we use a moving-threshold burst detection technique [13] that allows the detection of highlight on soccer game. This technique uses a moving-threshold that is obtained by computing the standard deviation and mean of the number of tweets during the playing of soccer games. For the implementation of this approach, we define a sliding window with the length $l$. We find that in our experiments the length $l$ equal to 15 seconds gives the best results. In the time sequence $(t_1, ...t_n)$ during the playing of soccer games, we have a sequence of sliding windows $(l_1, ...l_n)$. The sliding window $l_i$ at time $t_i$ contains $N(l_i)$ tweets. The algorithm is given below:

1) the length of the sliding window is $l$
2) At the time $t_i$, obtain the sliding windows $(l_1, ...l_i)$ and the corresponding numbers of tweets $N(l_1), ...N(l_i)$, and compute the values:
   - $mean_i = mean(N(l_1), ...N(l_i))$
   - $std_i = std(N(l_1), ...N(l_i))$
   - $MT_i = \alpha * (mean_i + x * std_i)$
3) The highlight in soccer games is detected in $l_i \;\; if \;\; N(l_i) > MT_i$

where $\alpha$ is the parameter used to relax the condition of moving-threshold and x is a constant defined between 1.5 and 2.0.Figure 6 shows the example of detecting highlights by applying a moving-threshold burst detection on the tweets posted online during the game playing between Brazil and Germany in the 2014 world cup.

### *B. Detecting Important Moments*

After identifying all the slopes that exceed the moving-threshold, we generate a list of spikes that correspond to the highlights in the soccer game. Then, each spike can be represented as tuple of $< Start - Time, Peak - Time, End - Time >$ (see figure 7). For each slope above the moving-threshold, we calculate the start time by finding backwards in time until we locate the point where the slope began increasing. The peak time is calculated by searching forward until we locate the point where the slopes start decreasing. we calculate the end time by searching forward from the peak time to locate the point where the slope begins to increase again. Given an important instant, as

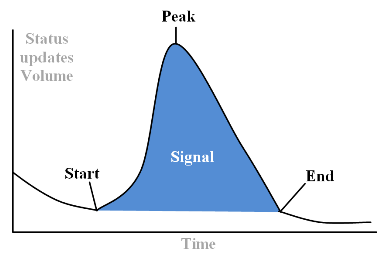


Figure 7. The skeleton of a spike, showing start-time, peak, and end-time

described by a spike tuple, we need to select a set of status that are performed in this instant. We could, choose to select all status updates sent between the start time and end time of the spike.Afterwards, we go to extract the semantics of the detected highlights from the tweets. We examine some tweets as follows: ”Brazil 2:0 Germany” (min. 23), ”Tooooooooooooooni kroos goooooooooool”, ”Germany goooooooooool”, ”Dante yellow card (min. 88)”, ”Brazil 1:7 Germany”. We can construct a highlight summary of soccer game playing by looking some keywords in those tweets such as ”Germany goooooooooool”, ”Dante yellow card (min. 88)” and the number of scores. Those tweets are collected during the soccer matches played between Brazil and Germany in the 2014 world cup. The first

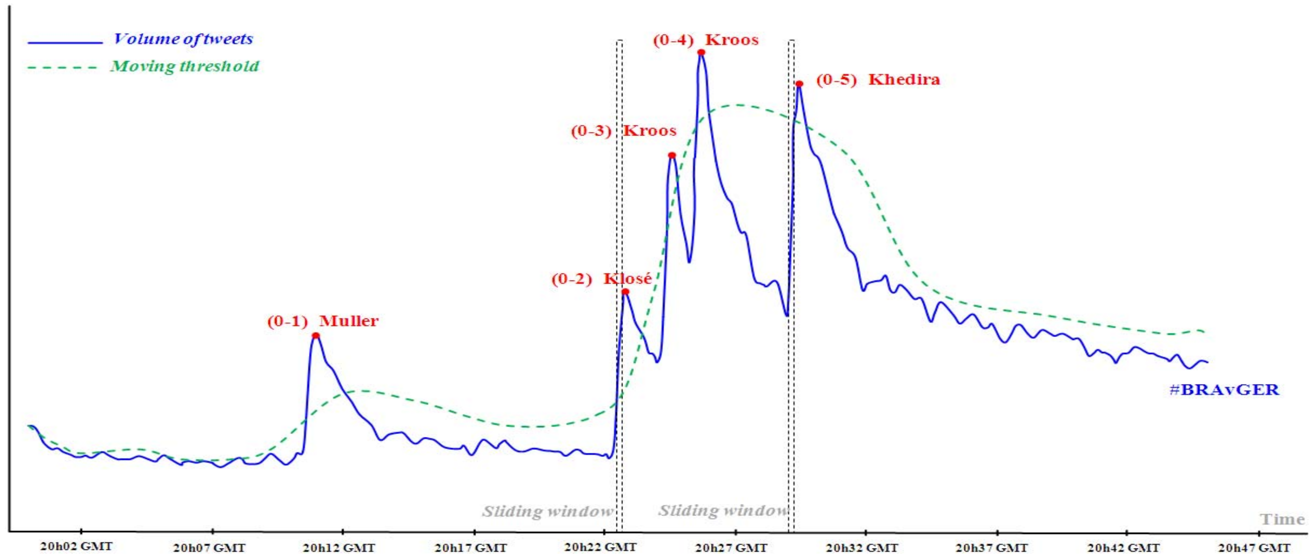


Figure 6. Highlights detection on soccer game by applying moving-threshold burst detection. This example taken from the 2014 World Cup between Brazil and Germany; the red dot: represents the detected highlights; Blue line: represents the number of tweets.

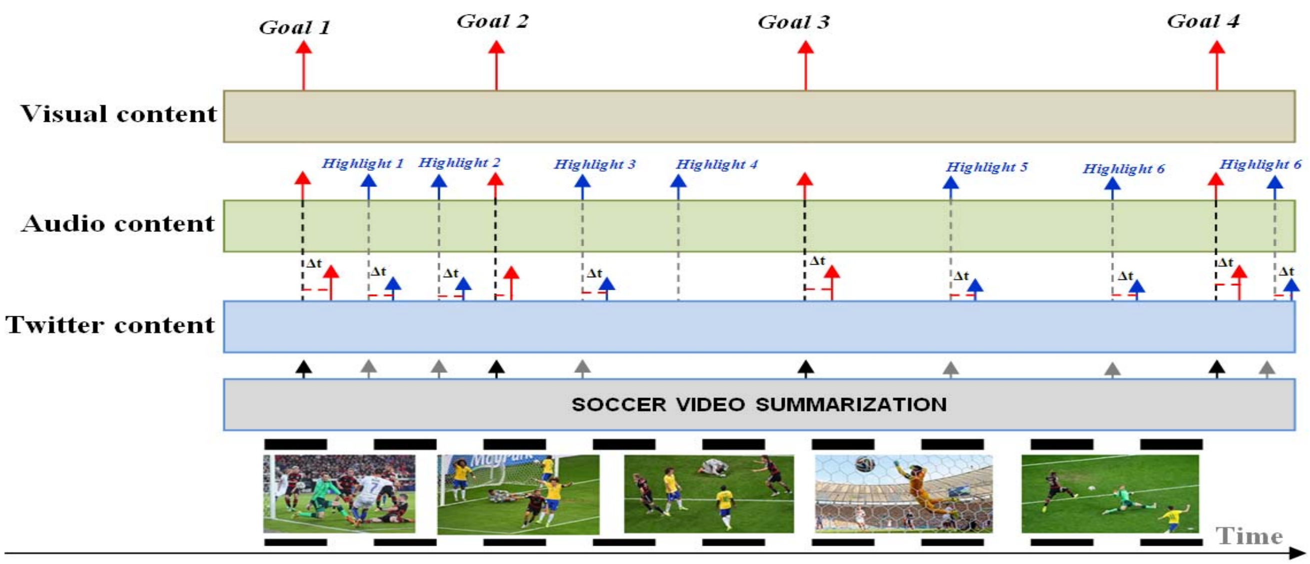


Figure 8. Highlights detection for video summarization

23 minutes of the game, Toni kroos from Germany scored a goal and made his team lead the game by 2-0. The semantic extraction step considers each tweet as a document and each detected highlight is a set of documents. In order to extract significant words for each highlight, we apply TF method [14] weighting approach to rank the words in each highlight. We measure the tf (term frequency) score of a word in a highlight as the number of times the word appears in the tweets of the highlight. We give some examples of words that can be considered as highlights: ”red card”, ”yellow card”, ”goal”, ”penalty”, ”cross-bar”, ”free-kick”, ”injured player”. We give in table 1 some statistics (Soccer matches played between Brazil and Germany in the 2014 world cup) relative to number of tweets per word that appears as highlight.

## IV. SOCCER VIDEO SUMMARIZATION USING HIGHLIGHTS

In this section, we will detail the process of fusion that is applied on all information obtained on the different stages of analysis. Which is the low-level processing represented by the analysis of visual content and the audio content performed on the video of soccer game; as well as the analysis of content of high-level represented by highlight

| Word appears in the tweets of the highlight | Number of tweets |
|---|---|
| Goal | 91250 |
| Yellow card | 5410 |
| Free-kick | 3125 |

Table I
NUMBER OF TWEETS PER HIGHLIGHT

| highlights detection | Precision |
|---|---|
| goal | 98% |
| penalty | 99% |
| red-cards | 89.7% |
| free-kick | 79.5% |
| corner | 72.4% |
| yellow cards | 71.2% |

Table II
PRECISION OF HIGHLIGHTS DETECTION

detection based on analyzing the tweets of Twitter (see figure 8). Red arrows that appear in figure 8 show the goals identified in the different part of the content (visual, audio, twitter). We find that the red arrows display on the part of the contents of twitter are offset with $\Delta_t$ compared to other red arrows. This discrepancy can be justified by the fact that tweets come after the scoring goals and since we have a percentage of 98% of goal detection in the part of analysis of visual content. Therefore we may determine with great precision the $\Delta_t$. This $\Delta_t$ will be used as a calibration tool between the audio content and twitter content. We use the content analysis of twitter to resolve the problem of highlights detection in the audio content cited in section II-C. To create the summary of the video, we use the detected goal in the analysis of visual content stage. Then we add the information derived from the analysis of the audio content and Twitter content to keep the most significant highlights.

## V. DATABASES

### A. Video database

to evaluate our approach we created a database of videos consisting of several soccer matches taken from the last World Cup. Our database contains over 30 hours of video a full-HD quality (1920*1080p), which represents 20 soccer games.

### B. Twitter Dataset

We build our tweet dataset by collecting the tweets posted during 20 soccer games of the 2014 World Cup. Tweets for this dataset were recorded through Twitters Streaming API using a track stream that receives tweets by using hashtag (#worldcup14) promoted by FIFA (www.fifa.com) and the other hashtag that have a relationship with the 2014 World Cup. We have collected more than 4 000 000 tweets related to the soccer matches of the 2014 World cup.

## VI. RESULTS AND DISCUSSION

In table II, we will show the results of highlights detection for video soccer summarization by using three methods based on: visual content analysis, audio content analysis and moving threshold burst detection on the tweets received from Twitter. To evaluate our approach, we compute the precision of our video soccer summarization. This is achieved by computing the highlights detection in soccer video matches. For each video soccer match we compute:

1) precision of goal detection
2) precision of penalty detection
3) precision of red-cards detection
4) precision of free-kick detection
5) precision of corner detection
6) precision of yellow cards detection

we give in the table II the precision results of highlights summarization obtained in 20 matches derived from the last World Cup. As you can see the precision percentage is higher than 98%. As regards the mean precision of soccer video summarization, the percentage reaches 84,96%.

## VII. CONCLUSION

In this paper, a new framework for summarization of soccer video has been presented. This framework allows soccer highlight detection for video summarization by using video content analysis and social media streams analysis (Twitter). Several methods have been proposed in this work, such as, the audio content analysis by using EMD, visuel content analysis and the algorithm of moving threshold burst detection used to detect the bursts of tweets on Twitter. The testing results proved that, the proposed framework is efficient with a mean precision of 84,96 %. From this results, we believe that the proposed approach can play an important role in improving the quality of the summarization of soccer video and the proposed framework can be applied to others sports, such as basketball, handball and baseball.

## REFERENCES


[1] Carling, Christopher, A. Mark Williams, and Thomas Reilly. Handbook of soccer match analysis: A systematic approach to improving performance. Psychology Press, 2005.

[2] Li, Baoxin, and M. Ibrahim Sezan. "Event detection and summarization in sports video." Content-Based Access of Image and Video Libraries, 2001.(CBAIVL 2001). IEEE Workshop on. IEEE, 2001.

[3] Ekin, Ahmet, A. Murat Tekalp, and Rajiv Mehrotra. "Automatic soccer video analysis and summarization." Image Processing, IEEE Transactions on 12.7 (2003): 796-807.

[4] Yow, Dennis, et al. "Analysis and presentation of soccer highlights from digital video." proc. ACCV. Vol. 95. 1995.

[5] Gong, Yihong, et al. "Automatic parsing of TV soccer programs. " Multimedia Computing and Systems, 1995., Proceedings of the International Conference on. IEEE, 1995.

[6] S. Intille and A. Bobick, ”Recognizing planned, multi-person action, ” Comput. Vis. Image Understand., vol. 81, no. 3, pp. 414445, Mar. 2001.

[7] V. Tovinkere and R. J. Qian, ”Detecting semantic events in soccer games: Toward a complete solution, ” in Proc. IEEE Int. Conf. Mult. Expo (ICME), Aug. 2001

[8] Xu, Changsheng, et al. ”Live sports event detection based on broadcast video and web-casting text.” Proceedings of the 14th annual ACM international conference on Multimedia. ACM, 2006.

[9] Nichols, Jeffrey, Jalal Mahmud, and Clemens Drews. ”Summarizing sporting events using twitter.” Proceedings of the 2012 ACM international conference on Intelligent User Interfaces. ACM, 2012.

[10] Zhao, Zhao, et al. ”Highlight summarization in sports video based on replay detection.” Multimedia and Expo, 2006 IEEE International Conference on. IEEE, 2006.

[11] Otsu, Nobuyuki. ”A threshold selection method from gray-level histograms.”Automatica 11.285-296 (1975): 23-27.

[12] Huang, Hai, and Jiaqiang Pan. ”Speech pitch determination based on Hilbert-Huang transform.” Signal Processing 86.4 (2006): 792-803.

[13] Gnther, David, HP. Seidel, and Tino Weinkauf. ”Extraction of dominant extremal structures in volumetric data using separatrix persistence.” Computer Graphics Forum. Vol. 31. No. 8. Blackwell Publishing Ltd, 2012.

[14] Hsieh, Liang-Chi, et al. ”Live semantic sport highlight detection based on analyzing tweets of twitter.” Multimedia and Expo (ICME), 2012 IEEE International Conference on. IEEE, 2012.