

Detection of Corner Event Based on Hidden Markov Model in Soccer Video

Haitao Yang¹, Jia Wang¹ and Jingmeng Sun^{2,*}

¹Beijing University of Technology, Beijing 100000, China

²Physical Education Department, Harbin Engineering University, Harbin 150000, China

yanghaitao@bjut.edu.cn

Abstract

A corner event detection method based on HMM (Hidden Markov Model) in soccer videos. Through analysis of the semantic structure of corner events, define and extract six multi-modal semantic clues to describe shot sequences, which constitute observation sequences as the HMM model input. By the iterative training of the HMM model and the continuous optimization of model parameters, construct the HMM model of corner events. From two aspects of audio and video, dig the inherent pattern of corner events and realize corner events automatic detection based on HMM accurately. The experimental results show that the present method achieves 89.6% recall and 96.3% precision, which has the better performance.

Keywords: Video semantic analysis, Hidden Markov Model, Audio and video features

1. Introduction

With the application and wide spread of sport videos, the automatic detection of highlight in those videos has become an important branch in the field of sport video retrieval area, receiving much attention from numerous researchers. In order to implement automatic detection of some specified events, scholars both home and abroad conducted intensive and systematical studies by applying extensively plenty of machine learning models. So far there're methods like dynamic Bayesian Network (DBN) [1-2], Bayesian Belief Network and Hidden Markov Model (HMM) [3]. HMM joins Markov chain model of temporal evolution and probability model of Bayesian network [4-5]. It's a broadly used statistical model. It's able to express the inherent uncertainties of observations in video episodes, which is more accordant with the temporal features of videos [6-7]. Also, it has effective expression form of video events and parameter learning algorithm [8-9]. The method based on HMM has been one of the most extensively discussed methods in the field of detecting sport video semantic events [10-11].

2. Retrieval of Semantic Clues of Corner Kick Episodes in Football Videos

In order to describe comprehensively and clearly corner kick episodes in football videos [12], according to the semantic content information covered by various shots of those episodes, we extract audio/video features from key frames of shots as to discover automatically from different aspects the inherent laws of the episodes, better depicting characteristics of different shots there. Hence it's necessary to use effective feature retrieval technique for the goal.

* Corresponding Author

2.1. Field Rate (FR)

Here we use HIS color model and cylindrical distance measure for field area segmentations. Make FR the field rate and Green Ratio (GR) the rate of green color in image frames. The quantitative rule is shown as (1):

$$FR = \begin{cases} 1 & GR(i) = \frac{D(i)}{\frac{2}{3} \times p \times q} > T_d \\ 0 & \text{else} \end{cases} \quad (1)$$

$GR(i)$ is grass green ratio of the frame i in camera. GR_{max} is the maximum ratio of the grass green. Figure 1 gives the field shot and non-field shot frames based on FR partition. In Figure 1 (a) and figure 1 (b), GR value is large, said rate of the grass green is the larger. The FR=1 field shot, figure1(c) is a non-field shot.



(a) field shot GR = 68.01% (b) field shot GR = 73.93% (c) (a) non-field shot GR = 8.74%

Figure 1. Semantic Shot Frame based on FR Partition

2.2. Player Ratio (PR)

The formula for calculating the aspect ratio and elongation of the minimum enclosing rectangle (MER) in player Aspect Ratio of MER (AR_MER) and Stretch Ratio of MER (SR_MER)(2)(3) for semantic shots are defined like (4):

$$AR_MER = \frac{W_MER}{H_MER} \quad (2)$$

$$SR_MER = \frac{\min(H_MER, W_MER)}{\max(H_MER, W_MER)} \quad (3)$$

$$PR = \begin{cases} 1, & AR_MER(i) < T_a, SR_MER(i) < T_h \\ 0, & \text{else} \end{cases} \quad (4)$$

Figure 2 shows the semantic shot representation frame. Figure2 (a) and figure 2(b) are close shot, and can clearly see the players on the court player body movement, players body accounted for a larger proportion. Figure2 (c) is a long shot.



Figure 2. Semantic Shot Frame based on FR Partition

2.3. Midfield Area (MA)

Midfield area shots have valuable information like kick-off, penalty kick, offence and defense. Episodes regarding corner kick and penalty kick occur in penalty area. If the kick-off circle is detected in the frame of one shot, the shot is called MA shot, without corner kick and penalty kick shots. In this case, MA-related features can be an important semantic clue for detecting non-corner kicks. Here we use Hough transformation to check out oval lines of the pitch. If such lines are found in frame images, the shot is called MA shot, *i.e.*, MA=1; otherwise MA=0. Figure 3(a) shows examples of representative frames of MA shots.

2.4. Penalty Area (PA)

Penalty area is a special zone in the football pitch. When a corner ball is kicked, there're global shots of both players striving fiercely in PA. But since only global shots have complete PA lines, we select key frames of global shots to detect if field lines exist in PA as thus to determine the penalty area. We employ Hough transformation method to match all straight lines in frame images. If two or more groups of parallel lines exist, the shot is PA shot, *i.e.*, PA=1; otherwise, it's non-penalty shot, *i.e.*, PA=0. Figure 3(b) presents examples of typical frames of PA shot.

2.5. Corner area (CA)

Generally when a player is serving a ball in the corner area, corner post and sideline intersects with base lines to form vertical angles in the shot. The feature is an important semantic hint to detect corner ball episodes. We utilize Hough conversion method to fit base lines and bottom lines in corner area, discerning the type of semantic shots by the angle β between two straight lines. The angle β is defined in the following (5):

$$\beta = \begin{cases} 1, & \begin{cases} 5^\circ < |\beta_1| < 85^\circ \\ 5^\circ < |\beta_2| < 85^\circ \\ 30^\circ < ||\beta_1| - |\beta_2|| < 85^\circ \end{cases} \\ 0, & \text{else} \end{cases} \quad (5)$$



Figure 3. Frame of Midfield Area Shot, Penalty Area Shot and Corner Area Shot in Soccer Video

2.6. Audio Energy

Audio features in football videos have rich semantic information, playing a supporting role in the semantic analysis of video structure. As seen in Figure 4(a), when a player is ready to service a ball, it is quiet in the field and accordingly AE is lower. In picture 4(b), footballers in both sides are struggling intensively in PA. When audiences are cheering together with narrator's excited voice, AE is higher. Therefore we use audio short-time energy feature $E(n)$ to make supplementary description of corner ball episodes, by the following calculation equation (6-7):

$$E(n) = \begin{cases} 1, & \frac{1}{N} \sum_{n=1}^N x^2(n) > \frac{1}{K} \sum_{k=1}^K x^2(k) \\ 0, & \text{else} \end{cases} \quad (6)$$

$$AE = \begin{cases} 1, & \sum_{m=1}^M E(m) > Th_i \\ 0, & \text{else} \end{cases} \quad (7)$$



Figure 4. Physical Shot of Soccer Video Corner Event

3. Detection of Corner Kick Episodes in Football Videos based on HMM

3.1. Markov Chains

The HMM model for corner ball episodes is designed to decide the structure of it, *i.e.*, the Markov chain shape of such episodes and the relationship between elements in videos and the model.

In sport videos, typical episodes occur by certain lens changing rules. To be specific, corner kick episodes contain local shots which show one player serves a ball in CA and global shots showing both players competing in the goal area. Learning that several scenes exist when a corner kick episode is happening, we get lens conversion way to which such an episode relates, and then determine and depict Markov link shape of the episode. Figure 5 gives the conversion chart.

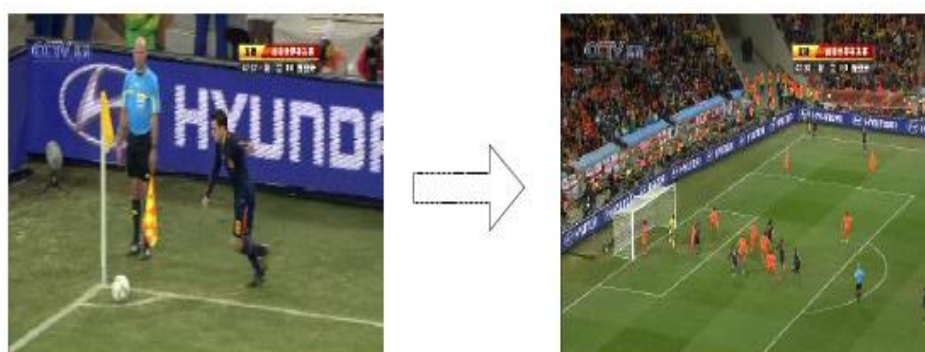


Figure 5. Transition Diagram of Corner Event Shot

3.2. Training and Reasoning of HMM

The detection of corner kick episodes is consisted of two parts:

(1) Regarding corner ball episodes, the HMM model is created, that is, to select enough video training samples for it. According to those samples, we can decide type of

state, number contained episodes; and sketch relative Markov chains. After pre-processing of samples in the training set, we determine the state descriptive method and complete the training process of HMM parameters.

(2) We use well-trained HMM model for the corner ball episodes to detect events in testing video fragments as to determine the classification of episodes. This part involves pre-processing of input video clips, further determination of parameters and the use of Viterbi algorithm to estimate state sequence.

HMM-based detection of corner ball episodes is implemented in the following steps:

Step 1 Build training data set

Select N_1 corner ball fragments and N_2 non-corner ball segments to form the training data set O of corner ball episodes; with regards to observation sequence $O_1, O_2, \dots, O_{N_1+N_2}$ of $N_1 + N_2$ shots in the set O , decide artificially the state of shots in each sequence, *i.e.*, global shot state or local shot state, to obtain the sequence $W_1, W_2, \dots, W_{N_1+N_2}$ of $N_1 + N_2$ states;

Step 2 HMM initialization

According to observation sequence $O_1, O_2, \dots, O_{N_1+N_2}$ of $N_1 + N_2$ shots in the set O and relative state sequence $W_1, W_2, \dots, W_{N_1+N_2}$, compute the initialized model parameter $\lambda = (\pi, A, B)$ of HMM model, where π is probability matrix of initial state; A is probability matrix of state transition; B is observation probability matrix;

Step 3 HMM training

As for training data set O , utilize Baum-Welch algorithm to perform iterative training of initialized model parameter $\lambda = (\pi, A, B)$; get the final optimized parameter $\bar{\lambda} = (\bar{\pi}, \bar{A}, \bar{B})$ of corner kick episode's HMM model, which is then built with the acquired optimal parameter;

Step 4 Build observation sequence

Perform physical shot segmentations of testing video clips; fetch six multimode semantic clues of key frames of every shot to constitute characteristic vectors; then such vectors of all shots in video clips form the observation sequence of the short video;

Step 5 Evaluation of HMM

After parameters in HMM model are defined, for input video clips pending for detecting, employ Viterbi algorithm to compute the possibility of the model generating state sequence.

4. Experimental Analysis and Results

The experiment data are chosen from 2010 South Africa FIFA, 2011 EPL and 2011 UEFA, MPEG format, 352×288 DPI, 25frames per second, Matlab R2008a. The experiment data for event detection is divided into training data and testing data, some used as training data parameters and some as testing data for validating the effect of the model. Of them, the training data of corner balls include 20 corner kick fragments and 10 non-corner kick fragments. Testing data have 30 corner kick clips and 20 non-corner kick clips. The effectiveness of the algorithm is measured by recall ratio and precision rate, defined as (8) (9):

$$Precision = \frac{N_c}{N_c + N_f} \times 100\% \quad (8)$$

$$Recall = \frac{N_c}{N_c + N_m} \times 100\% \quad (9)$$

4.1. Experiment on the Retrieval of Semantic Clues

4.1.1 Field rate Figure 6(a)(b)(c) is HIS histogram of training image and calculation results of the main color H_{mean} , S_{mean} and I_{mean} . Figure 6(g) also shows the results of segmenting the main color of video frame images by the cylindrical distance measure. Obviously, we realized well the separation of field and audience area, retained more detailed information and removed some pixel dots close to the main color in the audience area by filtering, such as crowds in green clothes and green flags which have impacts on feature extraction, avoiding incorrect and over segmentations. Table 1 gives the retrieval results of green field rate, which reveal that it realized effectively the extraction of field rate features, both recall and precision ratio reaching 100%.

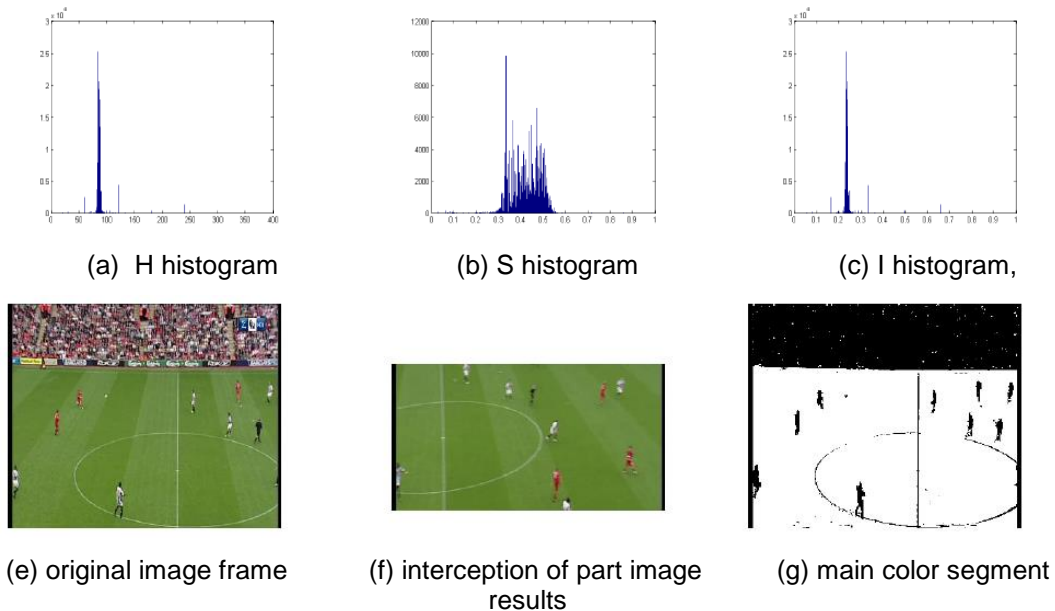


Figure 6. The Segmentation Results of Main Color and HSI Histogram of Training Images

4.1.2. Player rate: Figure 7 is the extraction result of field area and a binary image after morphological processing. From Figure 7 (c), we note that such processing padded well cavities caused during the segmentation. After burrs are eliminated, it's possible to form complete player connection area, laying the foundation for deciding the shape and size of subsequent candidate area. Table 1 lists the retrieval results of features regarding players' body rate, indicating that it extracted effectively such features, recall and precision ratio achieving respectively 84.64% and 95.65%.

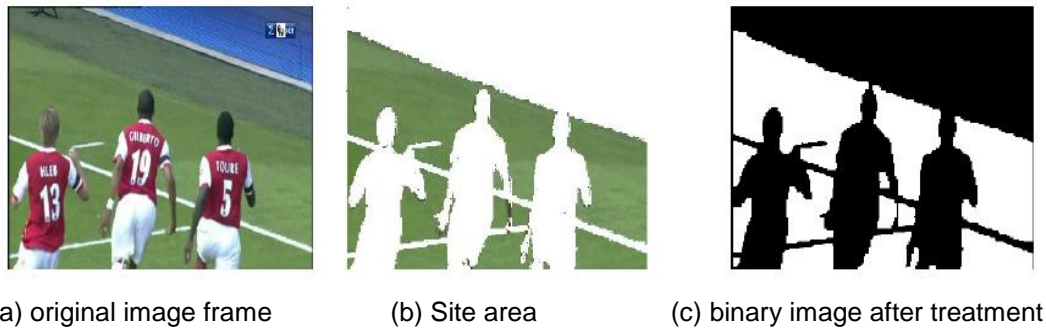
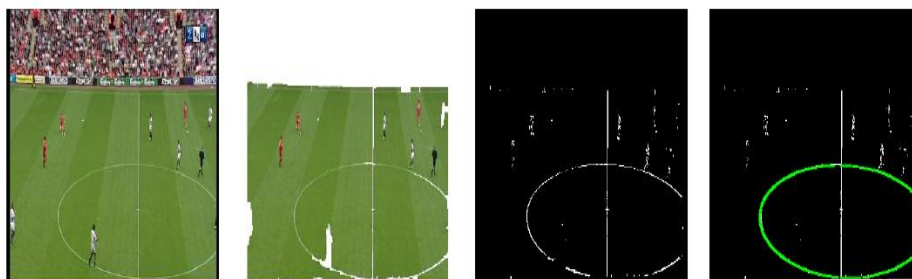


Figure 7. Extraction Results of Site Area

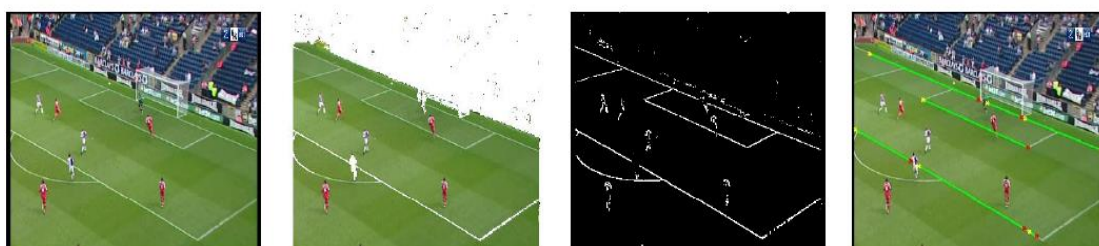
4.1.3. Oval Field Lines in Midfield Area: Figure 8(c) shows edge images detected with Canny operator, maintaining well detailed information of the ellipse. Green curves in the picture8 (d) are oval field lines detected through Hough transformation, matching with the actual oval field lines. As there're mutations with players and field background when edges are being detected, they may be regarded as edge lines and detected. During the sampling of Hough converted edges, there'll sampling points falling on players; and that it interferes with ellipse detection effect. Table1 presents the extraction results of oval field lines in midfield area, with recall and precision ratio at separately 83.61% and 91.07%, matching with the actual oval field lines. The detection effect is satisfactory, laying the foundation for the follow-up detection of corner kick episodes.



(a) original image frame (b) Site area (c) edge image (d) field lines extraction

Figure 8. Extraction Experiment Results of Elliptic Field Lines

4.1.4 Field Lines in Penalty Area: Figure 9 gives the result of extracting field lines in PA. Figure 9(c) shows edge images of using Canny operator to detect edges. Edges are clear, no cavities within edge lines and lower noise interference. The straight field lines in the figure9 (d) are fetched through Hough transformation. As seen, Hough transformation fits well the actual straight field lines in penalty area. Table1 is the retrieval results of field lines in PA, with recall ratio at 87.69% and precision rate at 93.44%. It implemented the retrieval of straight field lines in PA.



(a) original image frame (b) Site area (c) the edge image (d) Penalty area extraction

Figure 9. The Results of Field Lines Extraction of Penalty Area

4.1.5. Straight Field Lines in Corner Kick Area: Figure 10 (b) is the retrieval result in field area. By completely separating the field and background connection area, Figure 10 (c) drew up football pitch area. The Figure 10 (d) shows edge images of field lines which were detected. Those lines' edges are clear, helpful to the effective detection of subsequent field lines. They're linear field lines extracted through Hough transformation. Noticeably, it realized the extraction of linear field lines in corner kick area. Table 1 lists the retrieval results of field lines in corner kick area. It proved favorable detection effect thanks to extracting field lines in corner kick area.

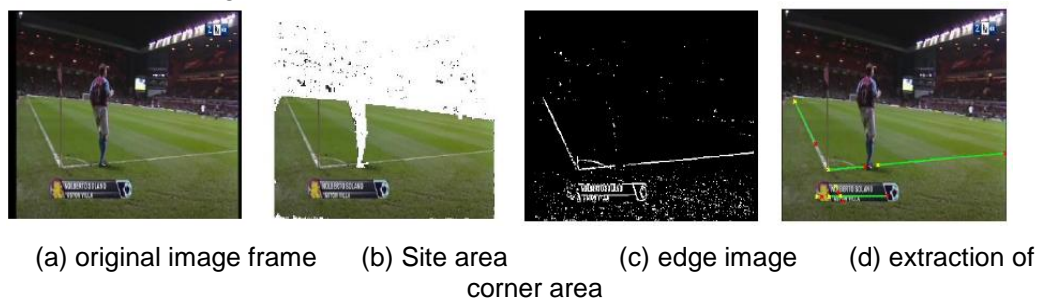


Figure 10. Line Field Extraction Results of Corner Area

4.1.6. Audio Energy: The experiment data have totally 45 video clips, including 30 wonderful fragments of 2010 South Africa FIFA and 15 not wonderful clips, wav format, lasting 60s, threshold $Th_i = 0.5350$. Figure 11 portrays one 60s-lasting wonderful voice clip signal $x(n)$ and energy diagram of each audio frame. There the peak values mean higher short-time energy values, *i.e.*, spectators cheering and narrator's louder voice, implying that audiences in the stadium are too excited and the match is fierce. Table1 is the retrieval results regarding audio features, suggesting that it can effectively check out fabulous voice fragments in football videos. The recall and precision rate reached respectively 81.08% and 88.24%. They can be used as significant auxiliary information to detect wonderful events.

4.2. Experiment on the Detection of Corner Kick Episodes based on HMM

Table1 gives experimental results of extracting semantic clues in six multiple modes. As observed, the recall and precision ratio of each retrieval test achieves above 80% on average. The detection effect proved good. By further experiment, we learnt that our defined multimode semantic retrieval method is effective. It is capable to dig out potential laws of corner kicks. It is simple and feasible, setting the foundation for detecting accurately and acquiring better detective performance.

In order to validate the proposed solution, we made contrastive analysis after the detection results were obtained. The experiment chose literature [13] as the control method. Table 2 lists the results of detecting corner kick episodes by our method and the control one. From the table2, the recall rate by our method reached 89.6% and precision rate 96.3%, increased respectively by 0.83% and 7.43% than the control method. The method for detecting corner kick episodes based on HMM improved remarkably in terms of both accuracy and recall rate, superior over the comparative method. Apparently, by defining and retrieving six multimode semantic clues to describe shots, building feature vectors and using them as observation inputs of the HMM model for corner kick episodes, the proposed method considered fully the semantic hierarchical structure of those episodes and realized well the automatic detection of them, with fused audio and video features enriching input field information of HMM model.

Table 1. Extraction Experiment Results of Multi-Pattern Semantic Cue

Semantic cue Extract type	Shot type	Actual number	Correct number	False number	Missing number	Precision%	Recall%
Field ratio	Field shot	57	57	0	0	100%	100%
Player ratio	Close shot	44	36	2	8	95.6%	83.6%
midfield zone field ground	midfield area shot	51	41	4	10	91.6%	84.6%
Field line of penalty area	penalty area shot	57	49	4	8	92.4%	85.9%
Corner area field lines	Corner zone shot	44	37	3	7	93.6%	86.2%
Audio features	Wonderful speech shot	30	21	4	7	88.2%	81.8%

Table 2. The Corner Event Detection Results with the Comparison Method

	Actual number	Correct number	False number	Missing number	Precision%	Recall%
In the paper	51	49	1	3	89.6%	96.3%
literature [13]	18	16	2	2	88.8%	88.9%

5. Conclusion

This paper presents a method for the detection of soccer video corner event based on HMM. Firstly, analyzed semantic structure of soccer video corner segment; secondly, on every shot key frame extracted six multimodal semantic cue. Audio and video features are fused, the description of the video shot, constitutes the observation value sequence. Finally, the optimal parameters are obtained by the iterative training model, and builds HMM model corner event, to achieve effective automatic detection based on HMM corner events, recall was 89.6%, precision was 96.3%.

Acknowledgment

This work was supported by The Fundamental Research Funds for the Central Universities. No. HEUCF151601.

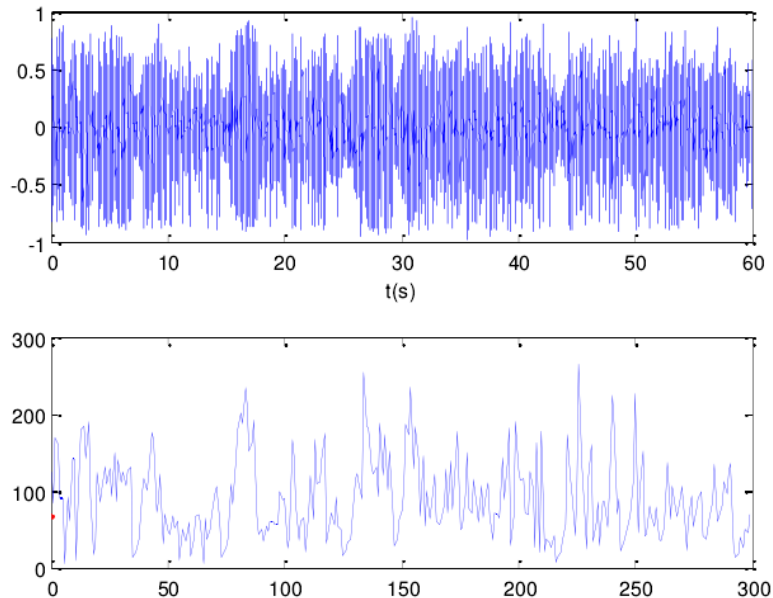


Figure 11. The Wonderful Speech Fragment $x(n)$ and Each Audio Frame Energy

References

- [1] L R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition", Proceedings of the IEEE, 1989, vol. 77, no. 2, pp. 257-286.
- [2] Ming, D. Liwei , M. J. Chan, "Soccer video highlights the fusion of HCRF and AAM detection.", Journal of computer research and development, vol. 1, (2014), pp. 225-236.
- [3] H. Shenghong, W. Baorong and J. Yufu, "Optimization strategy of adaptive video transmission based on content of lost frames", The application of computer system, vol. 7, (2014), pp. 100-105.
- [4] Hairihan, Research on tennis serve video event detection, Inner Mongolia Normal University, (2014).
- [5] Z. Yanjiao, "Regional map of target detection of video abstract Gauss", Hebei Normal University, (2014).
- [6] S. Chenhan, Methods and annotation of video structure extraction, Computer knowledge and technology, vol. 26, (2014), pp. 6178-6180.
- [7] Ming, D. Liwei and L. Yingying, "The soccer video highlights multidimensional semantic cues and HCRF model detection", Journal of Computer Aided Design & computer graphics, vol. 11, (2013), pp. 1715-1724.
- [8] Z. Yuan, "Analysis and research of tennis video semantic multi-modal fusion based", Nanjing University of Science and Technology, (2013).
- [9] Ming, D. Liwei and X. Wenjua, "Detection of soccer video semantic weighted and normalized score fusion rules", Journal of Computer Aided Design & computer graphics, vol. 2, (2013), pp. 167-174.
- [10] X. Faguo, Z. Xiang and L. Tao, "The tracking method for soccer video in HOG and particle filter based on ball", Electronic Science and technology, vol. 9, (2013), pp. 36-40.
- [11] D. Linna, "video annotation lens multi modal based on feature", Xi'an Electronic and Science University, (2013).
- [12] C. Yuting, "Traffic elements of perception and behavior understanding based on video image", Harbin Engineering University, (2013).
- [13] E. O. Rajae and O. H. T. Rachid, "Highlights recognition and learning in soccer video by using HMM and the Bayesian theorem", International Conference on Multimedia Computing and Systems, (2009), pp. 304-308.

Author



Haitao Yang, he received his B.S degree from Capital Institute of Physical Education, and received his M.S degree from Hebei Normal University. He is a lecturer in Beijing University of Technology. His research interests include Physical education and training.