

Sports Video Classification using Multi Scale Framework and Nearest Neighbor Classifier

S. Uma Maheswari^{1*} and R. Ramakrishnan²

¹Tamilnadu Physical Education and Sports University, Chennai, India; umasg76@yahoo.co.in

²Department of Advanced Sports Training & Technology, Tamilnadu Physical Education and Sports University, Chennai, India; ramkinr@yahoo.com

Abstract

Objectives: In order to achieve convenient sports video accessing without sequential scanning, automated sports video categorization is presented in this study. **Methods/Analysis:** In order to build efficient sports video categorizing system, edge features obtained from Non Subsampled Shearlet Transform (NSST) are taken into account. Then, sports genre categorization is done by Nearest Neighbor (NN) classifier due to its discriminative learning approach. The five sports category; tennis, cricket, volleyball, basketball and football are considered. **Findings:** To validate the proposed system based on NSST, experiments are carried using internal database video at frame level. Totally, 500 video clips are collected in which 100 video clips are gathered for each sports genre. The proposed system achieves maximum average classification accuracy of 94.80% at 4 directions of 2-scale NSST features while using city block distance measure in KNN classifier. For the same NSST features, the Euclidean, cosine and correlation distance measures gives an accuracy of 93.20%, 92.80% and 92% respectively. **Conclusion/Application:** The effectiveness of the system is clearly demonstrated by the experimental assessment. The proposed framework can adequately classify the sports video into one of the five predefined genre.

Keywords: Edge Features, Nearest Neighbor Classifier, NSST, Shearlet Transform, Sports Video Classification

1. Introduction

Computerized video classification has become more important due to vast availability of digital video contents based on contents in the videos. Among the various video categories such as cartoons, newspaper, commercial and sports, numerous researches have been carried out for sports video categorization due to its commercial demand. A new method of video summarization is presented in¹ based on motion analysis for sports video. Two optical flow algorithms are used to estimate the motion metrics. Then, different key frame selection criteria are used for each optical flow algorithms for video summarization and it is a threshold free approach.

Semantic event detection and classification using key frame selection approach is introduced in² for cricket videos. It consists of different stages such as hue histogram difference based key frames selection for indexing,

classification of frames into real time or replay based on logo transitions and then real time frames are classified as field view or non field view by colour features. Automated sports video categorization based on hierarchical decision making scheme is implemented for sports categories in³. The main decision making mechanism is a decision tree which generate hypotheses concerning the semantics of the sports video content. Hidden Markov Model (HMM) based decision making is employed.

Automated approach for personalized music sports video generation is described in⁴. Multimodal features are extracted from audio, video and text information. Dynamic programming is employed for the classification of personalized music sports video. Audio based sports video segmentation and detection is presented in⁵ for soccer video. At first, soccer sequence soundtrack is parameterized using Mel-Frequency Cepstral Coefficients (MFCC). Then it is segmented into homogenous components by

*Author for correspondence

using windowing algorithm with Bayesian model based decision process. Finally, series of HMM classifier is used for soccer video classification.

A new bag of words approach for sports video classification is implemented in⁶. Initially, Speeded Up Robust Features (SURF) are extracted from each frame. Then a codebook is generated using SURF by employing K-Means clustering algorithm. The histogram of codebook is computed and fed to Support Vector Machine (SVM) classifier for classification. Five layered hierarchical framework based sports video classification is explained in⁷. Top layer classification is performed by audio and video content analysis. Audio analysis is investigated by short time energy and zero crossing rate approaches. The video content is analyzed by HMM by extracting colour and motion features. Game specific rules are used by the lower layer to recognize major events of the game.

Statistical approach based shot segmentation and classification is presented in⁸. It segments and classifies the shots by same statistical inference features. Shot sequence probability is computed by Bigram using the relations between adjacent shots and feature sequence probability is dependent on inherent character of shot modeled by HMM. Edge feature based sports video classification is implemented in⁹ for five sports video categories. Edge direction histogram and edge intensity histogram are used as feature extraction technique. Auto associative neural network is employed as classifier, which classifies the sports category into one of the pre defined sports genre.

Video shot classification is described in¹⁰ for tennis. It is performed based on domain independent features and domain dependent features. By fusing these two features along with the domain knowledge, tennis video shots are classified into five classes. Multi modal approach based user generated mobile sports video classification system is designed in¹¹. Visual features, MFCC features and auxiliary features are used for classification by SVM.

Structure analysis of soccer video classification and segmentation based on HMM is presented in¹². The salient features; dominant color ratio and motion intensity in compressed domain are extracted. Then, a set of HMM is modeled for each state. Colour feature based sports video classification using HMM classifier is described in¹³. Three categories of sports genre such as hockey, football and golf are investigated. Color features are extracted from

the three sports category and trained by HMM classifier for classification.

Signature heat map based sports video classification is developed in¹⁴. Thermal imaging is used for detecting players and homography is used to detect their positions. Then heat maps are generated by summarizing Gaussian distributions representing people over 10-minute periods. Before classification, low dimensional heat maps are produced using fisher faces. Colour and motion features based sports news video shot classification is explained in¹⁵. Seven types of video shots and four types of sports video such as basketball, baseball, ice hockey and golf are examined. C4.5 decision tree classifier is employed for shot classification using 11-dimension shot feature.

Analysis between interest of playing traditional games and motion detection games are reviewed using ICT technique in¹⁷. The study has analyzed by playing interest of college students. Among two games, modern games play a consistency role among youngster. Cognitive radio based resource (channel) selection strategy is reviewed in¹⁸. A distributed, stochastic learning based resource selection and q-learning based resource selection is employed by the Secondary Users in a dynamic environment via game theoretic approach explicitly potential games.

In this paper, sports video classification using non subsampled shearlet transform and nearest neighbor classifier is presented. The remainder of this paper is organized as follows: The mathematical preliminaries used in the proposed sports video categorization system is explained in section 2. The implementation process of the proposed system is discussed in section 3. Section 4 describes the performance of the system and then conclusion is made in section 5.

2. Mathematical Preliminaries

The proposed sports video categorization system is constructed by using NSST and nearest neighbor classifier. This section describes the mathematical background of NSST and nearest neighbor classifier in detail.

2.1 Discrete Shearlet Transform

The multi-scale directional representation called the shearlet transform¹⁶ is employed in this study for feature extraction from the sports videos. An $N * N$ image consists of a finite sequence of values, $\{x[n_1, n_2]\}_{n_1, n_2=0}^{N-1, N-1}$ where

$N \in \mathbb{N}$. Identifying the domain with the finite group $\mathbb{Z}_N^2$, the inner product of image $x, y: \mathbb{Z}_N^2 \rightarrow \mathbb{C}$ is defined as

$$(x, y) = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} x(u, v) \overline{y(u, v)} \quad (1)$$

Thus the discrete analog of $L^2(\mathbb{R}^2)$ is $l^2\mathbb{Z}_N^2$. Given an image $f \in l^2(\mathbb{Z}_N^2)$, let $\hat{f}[k_1, k_2]$ denote its 2D Discrete Fourier Transform (DFT):

$$\hat{f}[k_1, k_2] = \frac{1}{N} \sum_{n_1, n_2=0}^{N-1} f[n_1, n_2] e^{-2\pi i (\frac{n_1}{N} k_1 + \frac{n_2}{N} k_2)} \quad (2)$$

The brackets in the equations $[\cdot, \cdot]$ denote arrays of indices, and parentheses $(\cdot, \cdot)$ denote function evaluations. Then the interpretation of the numbers $\hat{f}[k_1, k_2]$ as samples $\hat{f}[k_1, k_2] = \hat{f}(k_1, k_2)$ is given by the following equation from the trigonometric polynomial.

$$\hat{f}(\xi_1, \xi_2) = \sum_{n_1, n_2=0}^{N-1} f[n_1, n_2] e^{-2\pi i (\frac{n_1}{N} \xi_1 + \frac{n_2}{N} \xi_2)} \quad (3)$$

First, to compute

$$\hat{f}(\xi_1, \xi_2) \overline{V(2^{-2j} \xi_1, 2^{-2j} \xi_2)} \quad (4)$$

In the discrete domain, at the resolution level j , the Laplacian pyramid algorithm is implemented in the time domain. This will accomplish the multi scale partition by decomposing $f_a^{j-1}[n_1, n_2]$, $0 \leq n_1, n_2 < N_j - 1$, into a low pass filtered image $f_a^j[n_1, n_2]$, a quarter of the size of $f_a^{j-1}[n_1, n_2]$, and a high pass filtered image $f_d^j[n_1, n_2]$. Observe that the matrix $f_a^{j-1}[n_1, n_2]$ has size $N_j * N_j$, where $N_j = 2^{-2j} N$ and $f_a^0[n_1, n_2] = f[n_1, n_2]$ has size $N * N$. In particular,

$$\hat{f}_d^j(\xi_1, \xi_2) = \hat{f}(\xi_1, \xi_2) \overline{V(2^{-2j} \xi_1, 2^{-2j} \xi_2)} \quad (5)$$

Thus, $f_d^j[n_1, n_2]$ are the discrete samples of a function $f_d^j[x_1, x_2]$, whose Fourier transform is $\hat{f}_d^j(\xi_1, \xi_2)$.

In order to obtain the directional localization the DFT on the pseudo-polar grid is computed, and then one-dimensional band-pass filter is applied to the components of the signal with respect to this grid. More precisely, the definition of the pseudo-polar co ordinates $(u, v) \in \mathbb{R}^2$ as follows:

$$(u, v) = (\xi_1, \frac{\xi_2}{\xi_1}), \text{ if } (\xi_1, \xi_2) \in D_0 \quad (6)$$

$$(u, v) = (\xi_1, \frac{\xi_1}{\xi_2}), \text{ if } (\xi_1, \xi_2) \in D_1 \quad (7)$$

After performing this change of co ordinates, $\hat{g}_j(u, v) = \hat{f}_d^j(\xi_1, \xi_2)$ is obtained and for $l = 1 - 2^j, \dots, 2^j - 1$:

$$\begin{aligned} \hat{f}(\xi_1, \xi_2) &= \overline{V(2^{-2j} \xi_1, 2^{-2j} \xi_2) W_{jl}^{(d)}(\xi_1, \xi_2)} \\ &= g_j(u, v) \overline{W(2^j v - l)} \end{aligned} \quad (8)$$

This expression shows that the different directional components are obtained by simply translating the window function W . The discrete samples $g_j[n_1, n_2] = g_j(n_1, n_2)$ are the values of the DFT of $f_d^j[n_1, n_2]$ on a pseudo-polar grid. That is, the samples in the frequency domain are taken not on a Cartesian grid, but along lines across the origin at various slopes. This has been recently referred to as the pseudo-polar grid. One may obtain the discrete Frequency values of f_d^j on the pseudo-polar grid by direct extraction using the Fast Fourier Transform (FFT) with complexity $ON^2 \log N$ or by using the Pseudo-polar DFT (PDFT).

2.2 K-Nearest Neighbor Classifier

K-Nearest Neighbor (KNN) classifier is one of the simplest instance based learning algorithm. Based on distance function, KNN classifier assigns the class of the unknown object into one of the known training object's class. It requires training samples with class label. It calculates the distances between the unknown object and training samples. Then, it assigns the class label of the training samples which is nearest to the unknown object. The commonly used distance measures are Euclidean, city block, cosine and correlation. Table 1 shows the distance measures used in this study.

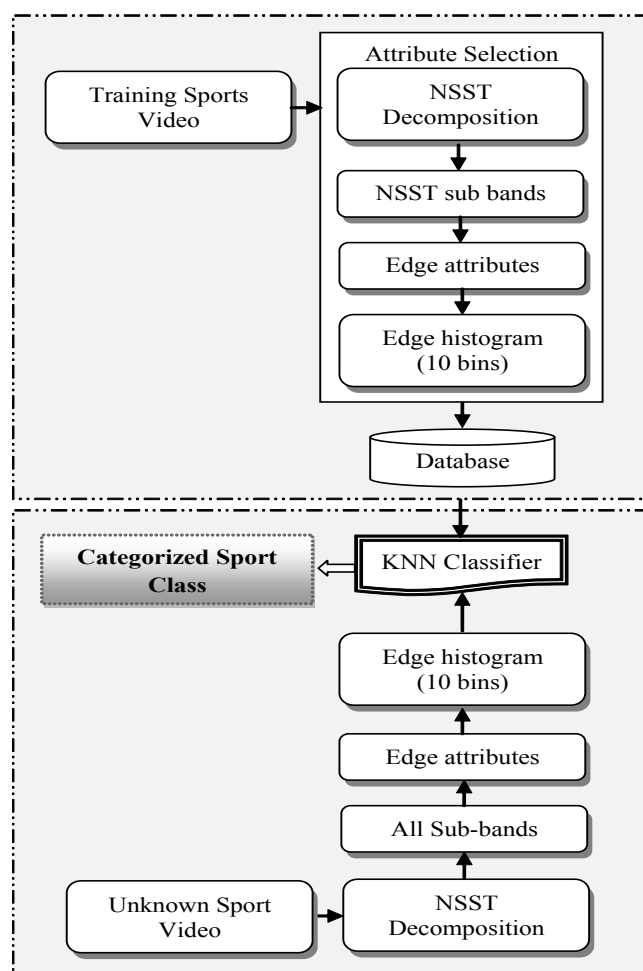
3. Sports Video Classification System Design

The proposed sports video classification system is analyzed by implementing two sub sequent functional building blocks: feature extraction or attribute selection and categorization. The proposed features are obtained from training videos at first functional block and the second block gets trained over the features extracted from first block for sports genre categorization. The schematic model of the proposed sports genre classification system is shown in Figure 1.

Table 1. Distance measures in KNN classifier

Distance measure	Formula
Euclidean	$D = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$
City block	$D = x_1 - y_1 + x_2 - y_2 + \dots + x_n - y_n $
Cosine	$D = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2} \sqrt{\sum_i y_i^2}}$
Correlation	$D = \frac{N \sum xy - (\sum x)(\sum y)}{\sqrt{[N \sum x^2 - (\sum x)^2][N \sum y^2 - (\sum y)^2]}}$

where $x = (x_1, x_2, x_3, \dots, x_n)$ and $y = (y_1, y_2, y_3, \dots, y_n)$ are n dimensional feature vectors

**Figure 1.** Schematic model of the proposed sports video categorization system.

3.1 NSST Based Feature Extraction

In any pattern recognition and machine learning approaches, feature extraction or attribute selection plays a crucial role, which tends to remove the redundant data and posses more intrinsic content of the original data. Hence, in order to design an efficient sport categorization system, novel multi-scale geometric characteristics of NSST is employed as feature extraction technique due to its optimal representation of image edges and capturing the geometric features of multidimensional data.

In order to extract edge features, NSST decomposition is carried out for each training video sequence frame by frame. Subsequently, NSST coefficients are obtained at different scales (2 to 3) and directions (2 to 64) that possess optimal distinctive edge attributes. However the dimension of the resultant attributes in each sub band has equivalent dimension of original data. It may certainly complicate the categorization system performance. To avoid this, NSST coefficients in each sub band is distributed into 10 bins in accordance with their magnitude ranges. This will reduce the high dimensional feature space of a video frame into only 10 attributes. The course of attribute extraction is repeated for all 500 video frames of a specific sports video and mean of extracted attributes are stored in database for further sports video categorization.

3.2 Categorization

The second functional block of the proposed system is categorization. Instance based KNN classifier is taken into account for unknown sport genre categorization by using similarity or distance function. At the start of the categorization stage, the edge strength attributes are obtained from the unknown (test) video sequence. Then, the test attributes and stored attribute database are fed into KNN classifier. It classifies the unknown sport genre into one of the predefined sports category by a majority vote of its computed neighbors using similarity measures. In this study, four distance measures named Euclidean, city block, correlation and cosine are analyzed.

4. Experimental Results

To validate the performance of the proposed sports video categorization system, an internal database is created. They are obtained by TV broadcast channels in different sessions. Totally, 500 video clips are collected in which 100 video clips are gathered for each sports genre;

volleyball, basketball, tennis, cricket and football. Each of the sports clip composed of 500 frames at the resolution of 128x128 with 20s duration. Random selections of 50 clips are used for training and remaining 50 clips are used for testing. The performance of the proposed sports genre classification system is analyzed in terms of classification accuracy. The classification accuracy is measured as the percentage of testing images categorized into the appropriate sports genre. Figure 2 shows the sample frames from the different videos in the database.

The performance of classifier not only depends on the features used for classification but also various parameters used in the classifier. Hence, to evaluate the system performance, the accuracies obtained by the proposed system using the following distance metrics such as Euclidean, city block, correlation and cosine are given in Table 2 to 5 respectively. It shows the classification accuracy of each sports genre and average classification accuracy.

It is observed from the Tables 2–5 that the proposed system achieves maximum average classification accuracy of 94.80% at 4 directions of 2-scale NSST features while using city block distance measure in KNN classifier. For the same NSST features, the Euclidean, cosine and correlation distance measures gives an accuracy of 93.20%, 92.80% and 92% respectively. While increasing the NSST decomposition from 2 to 3, the performance of the proposed sports video classification is degraded. It may be caused due to redundant information produced at higher level of decomposition. The maximum classification accuracy obtained by at 3rd level NSST decomposition using Euclidean, city block, cosine and correlation measure is 86.80%, 86%, 89.20%, and 85.60%

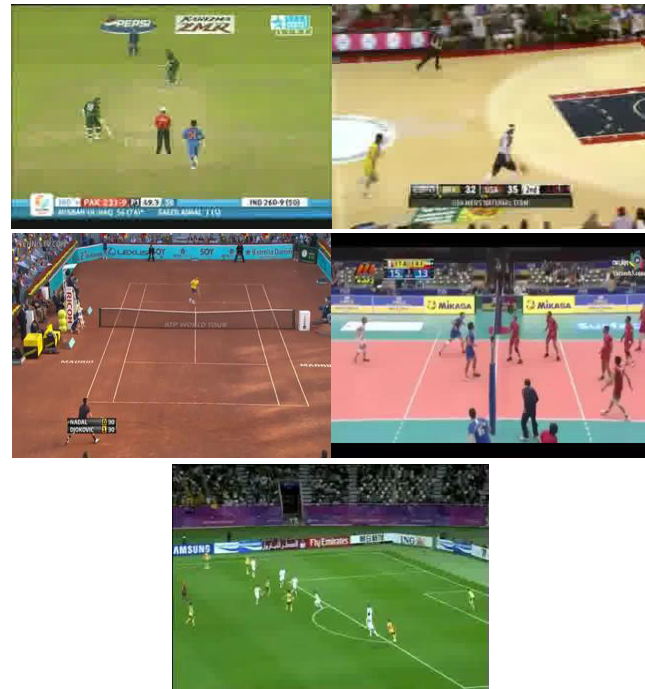


Figure 2. Sample frames from the different videos in the database.

respectively. Figure 3 shows the maximum average classification accuracy obtained by the proposed system.

5. Conclusion

In this paper, multi directional NSST and KNN classification based automated sports video classification system is presented. The proposed framework may surely helps in resourceful storage, speedy browsing and

Table 2. Accuracy of the proposed sports video classification system using NSST features along with Euclidean distance in KNN classifier

#NSST level	#NSST Directions	KNN distance measure - Euclidean					
		Classification Accuracy (%)					
		Cricket	Basketball	Tennis	Volleyball	Football	Average
2	2	76	94	92	94	86	88.40
	4	84	100	94	96	92	93.20
	8	84	100	94	100	90	93.60
	16	84	100	94	98	88	92.80
	32	86	100	92	98	90	93.20
	64	78	94	98	96	82	89.60
3	2	78	100	96	90	70	86.80
	4	72	98	94	96	70	86
	8	70	98	92	94	68	84.40
	16	72	94	86	94	62	81.60
	32	72	98	94	96	66	85.20
	64	74	96	82	94	76	84.40

Table 3. Accuracy of the proposed sports video classification system using NSST features along with city block distance in KNN classifier

#NSST level	#NSST Directions	KNN distance measure - city block					
		Classification Accuracy (%)					
		Cricket	Basketball	Tennis	Volleyball	Football	Average
2	2	76	98	94	98	88	90.80
	4	86	98	94	98	98	94.80
	8	86	98	94	98	94	94
	16	86	98	94	100	92	94
	32	86	98	94	98	92	93.60
	64	84	96	98	96	88	92.40
3	2	76	98	98	90	66	85.60
	4	74	98	96	94	68	86
	8	72	98	96	94	68	85.60
	16	74	92	92	96	60	82.80
	32	72	98	94	96	66	85.20
	64	74	90	84	94	74	83.20

Table 4. Accuracy of the proposed sports video classification system using NSST features along with cosine distance in KNN classifier

#NSST level	#NSST Directions	KNN distance measure - cosine					
		Classification Accuracy (%)					
		Cricket	Basketball	Tennis	Volleyball	Football	Average
2	2	66	94	94	96	82	86.40
	4	80	98	96	96	94	92.80
	8	76	98	96	96	90	91.20
	16	80	98	96	96	90	92
	32	78	98	92	96	90	90.80
	64	72	92	98	96	72	86
3	2	78	98	96	94	68	86.80
	4	68	98	94	98	70	85.60
	8	70	98	92	96	68	84.80
	16	64	92	88	96	62	80.40
	32	76	100	90	98	82	89.20
	64	70	98	80	96	72	83.20

Table 5. Accuracy of the proposed sports video classification system using NSST features along with correlation distance in KNN classifier

#NSST level	#NSST Directions	KNN distance measure - correlation					
		Classification Accuracy (%)					
		Cricket	Basketball	Tennis	Volleyball	Football	Average
2	2	62	92	94	94	80	84.40
	4	80	98	96	94	92	92
	8	76	98	96	96	90	91.20
	16	76	98	96	96	90	91.20
	32	72	98	92	94	90	89.20
	64	66	94	98	96	74	85.60
3	2	68	96	96	94	64	83.60
	4	64	96	94	98	64	83.20
	8	66	96	92	96	64	82.80
	16	62	90	90	94	62	79.60
	32	72	96	84	96	80	85.60
	64	66	98	80	94	64	80.40

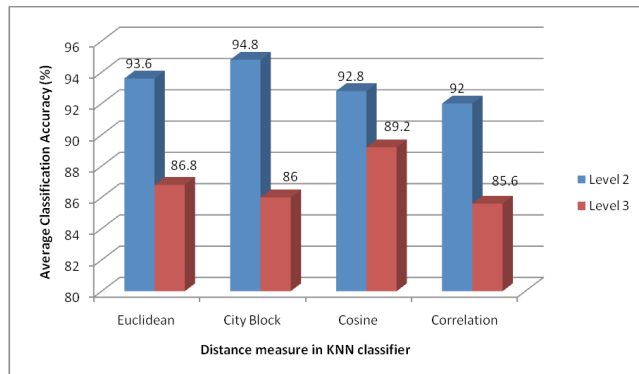


Figure 3. Maximum average classification accuracy obtained by the proposed system.

retrieval of huge collection of sports video data without losing important characteristics. To facilitate the proposed sports video categorization system, histogram of edge features is extracted from the video sequences by NSST decomposition at various level and directions. Then, the query sports video is categorized using KNN classifier into one of the predefined five sports genre. The experimental results show that the proposed sports genre classification system produces 94.80% at 2nd level decomposition with 4 directions of NSST.

6. References

- Mendi E, Clemente HB, Bayrak C. Sports video summarization based on motion analysis. *Comput Electr Eng*. 2013; 39(3):790–6.
- Goyani MM, Dutta SK, Raj P. Key frame detection based semantic event detection and classification using hierarchical approach for cricket sport video indexing. *Advances in Computer Science and Information Technology*. Springer: Berlin Heidelberg. 2011; 131:388–97. doi: 10.1007/978-3-642-17857-3_39.
- Jaser E, Kittler J, Christmas W. Hierarchical decision making scheme for sports video categorisation with temporal post-processing. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2004; 2:908–13. doi: 10.1109/CVPR.2004.1315262.
- Wang J, Cheng E, Xu C, Lu H, Tian Q. Generation of personalized music sports video using multimodal cues. *IEEE Trans Multimed*. 2007; 9(3):576–88. doi: 10.1109/TMM.2006.888013.
- Baillie M, Jose JM. An audio-based sports video segmentation and event detection algorithm. *IEEE Conference on Computer Vision and Pattern Recognition Workshop*; 2004. p. 110. doi: 10.1109/CVPR.2004.24.
- Duong D, Dinh TB, Dinh T, Duong D. Sports video classification using bag of words model. *Intelligent Information and Database Systems*. Springer: Berlin Heidelberg; 2012. p. 316–25. doi: 10.1007/978-3-642-28493-9_34.
- Kolekar MH, Sengupta S. A hierarchical framework for generic sports video classification. *Computer Vision*. Springer: Berlin Heidelberg; 2006. p. 633–42. doi: 10.1007/11612704_63.
- Yang Y, Lin S, Zhang Y, Tang S. Statistical framework for shot segmentation and classification in sports video. *Computer Vision*. Springer: Berlin Heidelberg; 2007. p. 106–15. doi: 10.1007/978-3-540-76390-1_11.
- Mohan CK, Yegnanarayana B. Classification of sport videos using edge-based features and auto associative neural network models. *Signal, Image Video Process*. 2010; 4(1):61–73. doi: 10.1007/s11760-008-0097-9.
- Yu XD, Duan LY, Tian Q. Shot classification of sports video based on features in motion vector field. *Advan Multimed Informat Process*. Springer: Berlin Heidelberg; 2002. p. 253–60. doi: 10.1007/3-540-36228-2_32.
- Cricri F, Roininen M, Mate S, Leppanen J, Curcio ID, Gabbouj M. Multi-sensor fusion for sport genre classification of user generated mobile videos. *IEEE International Conference on Multimedia and Expo*; 2013. p. 1–6. doi: 10.1109/ICME.2013.6607536.
- Xie L, Xu P, Chang SF, Divakaran A, Sun H. Structure analysis of soccer video with domain knowledge and hidden Markov models. *Pattern Recogn Lett*. 2004; 25(7):767–75. doi: 10.1016/j.patrec.2004.01.005.
- Hanna J, Patlar F, Akbulut A, Mendi E, Bayrak C. HMM based classification of sports videos using color feature, *IEEE International Conference on Intelligent Systems (IS)*; 2012. p. 388–90. doi: 10.1109/IS.2012.6335247.
- Gade R, Moeslund TB. Sports type classification using signature heat maps. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*; 2013. p. 999–1004. doi: 10.1109/CVPRW.2013.145.
- Wang DH, Tian Q, Gao S, Sung WK. News sports video shot classification with sports play field and motion features. *IEEE International Conference on Image Processing*; 2004. 4:2247–50.
- Easley G, Labate D, Lim WQ. Sparse directional image representations using the discrete shearlet transform. *Appl Computat Harmonic Analys*. 2008; 25(1):25–46. doi: 10.1016/j.acha.2007.09.003.
- Kannan M, Geetha M. An analysis between traditional and motion detection game-using ICT techniques. *Indian J Sci Technol*. 2014; 7(12):1956–62.
- Poongodi K, Singh HK, Kumar D. Co-Operation based resource selection in cognitive radio network via potential games. *Indian J Sci Technol*. 2015; 8(S2):63–9.