Audience Shot Detection for Automatic Analysis of Soccer Sports Videos

Kazimierz Choroś[a]
Department of Applied Informatics, Wrocław University of Science and Technology, Wyb. S. Wyspiańskiego 27, Wrocław, Poland

Keywords: Content-based Video Indexing, TV Sports News, Soccer Sports Videos, Sports Videos Categorization, Video Shot Categorization, Sports Discipline Recognition, Audience Shots, Dominant Color, Segment Color Histograms, Shots Detection, Face Detection.

Abstract: The automatic categorization is all the time a great challenge of content-based indexing of sports videos. The great part of different video archives, portals, Web databases contains a huge amount of sports videos data. One of the most significant processes of sports news videos analysis is automatic recognition of a sports discipline reported in a video. Different strategies are applied: pattern frame comparison, line detection in playing fields, player detection, sports equipment detection, or detection of superimposed text, and others. Usually audience shots are processed like other non-player shots, considered as not useful for video content analysis. This paper presents an approach of automatic detection of audience shots which are however useful for automatic categorization of sports videos. The audience shots in sports videos can be considered as very informative parts helping to detect and recognize not only sports disciplines, but also nationality or club membership, as well as emotions of supporters. The method is based on the integration of the analysis of segment color histograms of video frames, detection of shots, and face detection. Color histograms are applied to detect audience frames and shots. Because the dominant color as a unique criterion is not efficient in audience detection this procedure has been improved by analyzing not only single frames but sequences of frames belonging to one shot. Then a face detection method has been introduced to find the most suitable audience shots for content analysis of sports videos. The tests have been performed on soccer sports videos.

1 INTRODUCTION

The sports videos are very popular and probably the most frequently viewed in the Internet. The extremely huge amount of videos being uploaded every day to different video data bases, archives, Internet portals, etc. generate the necessity to automatically index videos basing on theirs contents. However, the efficiency of automatic content-based video indexing is still not satisfactory. Only for some special kinds of videos different approaches proposed during last decades seem to be promising. For example sports disciplines of sports reports in TV sports news videos are automatically recognized on the basis of content-based analysis and then they are used to index the sports news videos.

The content can be recognized using different strategies such as pattern frame comparison, line detection in playing fields, player detection, sports equipment detection, or detection of superimposed text, and others (Assfalg et al., 2002; Bertini et al., 2005; Choroś, 2016; Shih, 2017). There are more and more commercial applications (Thomas at al., 2017).

The sports videos of a given sports discipline can be then easier retrieved in video archives. The retrieval systems are more efficient and the retrieval procedure can be more sophisticated when weighted indexing methods are applied (Choroś, 2016).

Audience shots are usually omitted during the content-based analysis of sports news because they do not present player games although they are parts of sport events. They are similarly treated as publicity, breaks in the game, any organizational breaks due to wheather or technical problems. During some soccer championships the television editors avoid showing the audience, in consequence the audience segments are not very frequent. But in others mainly after scoring the goal and the final referee's whistle the

[a] https://orcid.org/0000-0001-6969-976X

Choroś, K.
Audience Shot Detection for Automatic Analysis of Soccer Sports Videos.
DOI: 10.5220/0010404806880695
In *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2021) - Volume 5: VISAPP*, pages 688-695
ISBN: 978-989-758-488-6

reaction of the audience – joy or sadness – can be observed on the TV screen.

Our thesis is that such parts of the sports videos carry very valuable information which can be used in the content-based analysis of sports videos.

What is the information we could get when analyzing the audience shots? Audience shots can help to automatically detect and recognize:

- sports discipline – the audience for some sports disciplines is very specific because of specific clothes, specific supporters accessories, banners and flags with the inscriptions of supporter cities, names of athletes, names of fan clubs, etc.;
- nationality of player teams – during the great world cup, Olympic Games, or regional European or South American championships the soccer fans wear the clothes in national colors;
- name of the soccer club – usually in the special audience sector of avid fans most of people are dressed in clothes in club colors, sometimes with the names of their favorite players;
- emotions of supporters – recognition of people emotions is very useful for the detection for example of highlights such as scoring goals, nice kicks, fouls, penalties, etc.

The conclusion is that during the content-based analysis of sports news audience shots should not be omitted similarly as all other non-player shots. They can help us to get useful information. But the problem arising is how to detect audience shots in sports videos and which audience shots can be useful for specific procedures of content analysis, specific for a given kind of sports shots. The approach presented in this paper proposes an efficient solution.

2 RELATED WORK ON AUDIENCE ANALYSIS

Detection of audience shots has been discussed and tested in many scientific and experimental studies. In early research described in (Assfalg et al., 2002) it was observed that the audience scenes are typically characterized by more or less uniform distributions for edge intensities, segments orientation, and hue. Moreover, it was argued that the color distributions of the audience shots are much more uniform than playing field or player shots.

In the other experimental studies the audience shots, coach shots, and also other shots were denoted as out of field shots (Ekin et al., 2003). Furthermore, it was observed that an out of field shots of the audience as well as close-up shots of a player often indicate a break in the game. This was the reason that close-up shots and out of field shots were classified in the same category due to their similar semantic meaning.

The similar procedures can be applied for baseball videos because a baseball playing field is similar to soccer sports terrains. In (Kuo et al., 2011) baseball shots were classified into five most popular types such as pitch, infield, outfield, close-up and audience using template datasets. It was proposed to examine color histogram feature of non-field close-up and audience shots but by extracting the color distribution from the specified region, chosen in the central part of the analyzed frames.

Also in other experimental research non-court view shots were as usually composed of player close up shots and audience shots. In (Zhang et al., 2011) the analysis of tennis videos was based on court view shots which contains full court lines, because as it was stated, court view shots in tennis videos have stable line character which doesn't exist in non-court view shots, so also in audience view shots. Lines in tennis videos are very significant. Tennis shots can be selected from TV sports news basing on the minimum set of lines sufficient to detect a tennis court (Choroś, 2012). Not all lines need to be detected. The most typical for tennis court are two pairs of long vertical lines but due to the perspective view converging to the top of the image in a TV broadcast and then one long horizontal line at the bottom of the image. Such minimum set of lines is sufficient for the categorization of tennis shots. However, lines are not present in the audience shots. In (Jiang et al., 2011) tennis videos were also analyzed. There it was observed that audience shots have a high similarity to close-up shots. Whereas the long view audience shots are characterized by the highest edge density.

The discrimination between audience shots and close-up shots could be achieved by color analysis and edge information extraction. Close-up shots are dominated by a large ratio of skin pixels, while its edge density is usually lower than that of long view audience shots (Fang et al., 2013).

In hockey puck detection and tracking systems applied for the detection of video highlights in ice-hockey videos the sequences in which the puck is not visible for a long time are excluded (Yakut et al., 2016). Usually the camera is pointing to the audience, or fights between players. So, in such systems audience shots are simply excluded.

Audience shots were identified not only in sports videos but also for example in learning media (Li et al., 2005). Six types of frames were defined: slide, web-page, instructor, audience, picture-in-picture, and miscellaneous. Those shots were treated as the audience shots which contained a long frame sequences of the meeting room. Whereas in (Daudpota et al., 2019) three to six types of shots were defined, three mandatory shots: shots of host, shot of one or more guests, a combined shot of host and guests, but also in addition shots of all audience, shot of someone from the audience, and shot of environment to give an idea to viewers about the location of the show. Clustering-based shot detection algorithm was used for the shot detection.

However, in all these approaches the audience shots were treated as out of field shots, so, in the same way as other close up shots presenting coaches, referees, players, etc. In general in the color-distribution approaches close up shots and audience shots were not discriminated. Whereas in our research the goal was to detect audience shots, mainly medium shots with audience, such shots that can help to improve the results of video content analyses. As it was observed such shots can help to automatically recognize the sports category, name of sports club, place of sports competitions, amazing highlights of the sports events, etc. Usually the best shots for content analysis are player shots, mainly long shot displaying the global view of the field. Non-player shots seem to be useless. Whereas in this paper audience shots are seen as shots valuable for content based analysis.

3 AUDIENCE FRAMES

Sometimes the audience is multicolored (Figure 1) mainly in long view shots and not useful for recognizing the nationality or the club of the soccer team. Long view audience shots present a large number of spectators, it normally results in a large amount of edge information. However, during the soccer matches after scoring the goal the most frequently the joy of the most avid fans gathered in one of the special sector of the soccer stadium is shown. The great number of soccer fans are very frequently wearing shirts in typical national colors (Figure 2–5). The analysis of colors in medium shots with audience can suggest the country whose team is playing the match or the soccer club.

So, the content analysis of broadcasted sports news can result not only in recognizing the sports discipline but when the soccer shots are detected the analysis of audience shots can suggest which country or club team plays a match. Of course in some cases the ambiguities may occur. The red color is typical for fun wears not only from Switzerland but also from Croatia, Denmark, Poland, or others.

Figure 1: Multicolored audience in a long view shot.

Figure 2: Yellow Brazilian audience.

Figure 3: White and red Polish audience.

Figure 4: White and blue stripes of Argentinean audience.

Figure 5: Orange Dutch audience.

Figure 6: Long view field shot with players.

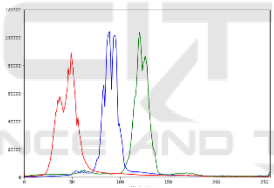

Figure 7: RGB histogram of long view field shot with players.

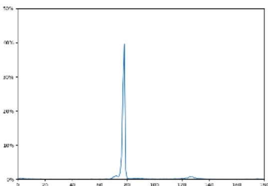

Figure 8: Hue histogram of long view field shot with players.

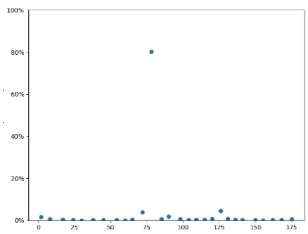

Figure 9: Hue segment histogram of long view field shot with players.

4 HISTOGRAMS OF DIFFERENT FRAME TYPES

The most typical view for soccer matches is a long view of soccer field (Figure 6). It was noticed in many other studies that in such video frames the green color is dominant (Figures 7–8). So, the dominant color in video frames was a discriminate feature.

In this analysis segment histograms were used to reduce the influence of small color differences, natural in real videos (Figure 9). In segment histograms the color space is divided into fixed blocks (segments). Segments were defined starting with the most frequent hue values by assigning other colors with hue values differing by no more than five from the main hue value in the segment.

The audience shots have another characteristic. Usually we do not observe a dominant color in frame histograms (Figure 10). Even when the frame is not reach in colors its histogram is relatively flat (Figures 11–13).

The question arising is what kind of shots with audience is the most suitable for automatic content analysis. Which audience shots can help to automatically recognize these aspects discussed in the introduction, i.e. sports discipline, nationality, sports club, emotions, etc. The long view audience shots which are relatively easy to distinguish among playing fields shots are not those we are looking for. Individuals are very small objects, so the detection of their clothes, faces, gestures is not possible.

Figure 10: Audience frame without the dominant color.

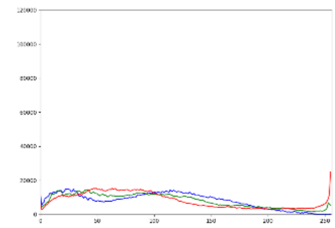

Figure 11: RGB histogram of an audience shot.

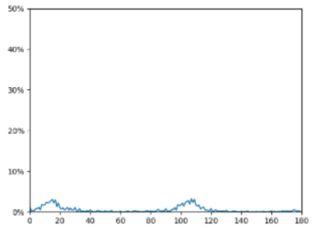

Figure 12: Hue histogram of an audience shot.

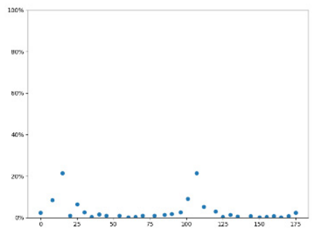

Figure 13: Segment histogram of an audience shot.

On the other hand the very close-up shots are not also useful. The face of a person on the whole screen can belong to a player, referee, coach, or spectator. It seems that the best audience frames for the content analysis are medium frames (Figure 14).

Figure 14: Examples of audience frames in soccer videos. In all frames the audience is seen but the most useful frame for the audience analysis is the frame in the upper right corner – this frame is selected from the medium view shot.

This is the reason the proposed approach is based not only on the analysis of color histograms, but it takes also into account the neighboring frames – video is not a single image, and also the number of faces detected in the analyzed frames. It was observed that when audience was presented the camera was very stable comparing to the playing field shots.

5 AUDIENCE SHOT DETECION

Three different soccer videos were used in the tests (Table 1). In these three sample movies the audience was presented a few times. First the segment color histograms were analyzed, then audience shots extracted, and final improvement consisted in analyzing the number of faces in frames.

The standard efficiency measures were applied to analyze and to compare the experimental results: recall, precision, fallout, sensitivity, an F-measure.

These experimental studies were conducted as part of the AVI project (Choroś, 2010).

Table 1: Videos used in the experiments.

Video	Total number of frames	Total number of playing field frames	Total number of audience frames
Video 1	5412	2063	275
Video 2	6001	2110	569
Video 3	6751	2511	562

5.1 Detection based on Green Color of Soccer Field

The dominant green color of soccer playing field is well-known feature used to detect soccer player shots in sport videos. But it can be also used to eliminate such shots when audience shot are being detected. In our experimental studies we tested by examining the segment color histograms what is the minimum share of green color in a frame to assume that this frame is a soccer field frame.

Table 2: Detection of playing fields shots.

Minimum share [in %] of green color in a frame	Recall	Preci-sion	Fallout	Sensi-tivity	F-Measure
10	1.00	0.43	0.77	0.51	0.60
20	1.00	0.46	0.69	0.57	0.63
30	1.00	0.49	0.62	0.61	0.66
40	1.00	0.55	0.49	0.69	0.71
50	1.00	0.60	0.40	0.74	0.75
60	1.00	0.66	0.32	0.80	0.79
70	1.00	0.87	0.09	0.95	0.93
75	0.81	1.00	0.00	0.93	0.89
80	0.54	1.00	0.00	0.83	0.69
85	0.32	1.00	0.00	0.75	0.49

The best threshold was 70%, that is if the green color is dominant for 70–75% of pixels in the frame it is very probable it is a playing field frame (Table 2). Then recall was equal to 1, precision – 0.87, and F-measure reached the best value 0.93.

If the desirable audience views are medium views they have not the green playfield as a background. The dominant green color can be then a criterion for rejection of these frames as they do not present the audience. Several values of the share of green color were tested (Table 3). The results showed that for audience frames the share of the green color in a frame was not greater than 10%. However, such a criterion did not ensure the high values of precision, only 0.59. Nevertheless, it means that it permitted to reduce the number of frames for further analysis to less than half.

5.2 Selection of Audience Frames based on Green Color of Soccer Field

Table 3: Selection of audience frames using segment color histograms.

Maximum share [in %] of green color in a frame	Recall	Preci-sion	Fallout	Sensi-tivity	F-Measure
5	0.95	0.68	0.04	0.95	0.76
10	1.00	0.59	0.07	0.93	0.70
15	1.00	0.49	0.11	0.90	0.63
20	1.00	0.44	0.13	0.88	0.58
25	1.00	0.39	0.15	0.86	0.54
30	1.00	0.36	0.17	0.84	0.51
35	1.00	0.31	0.21	0.80	0.46
40	1.00	0.26	0.26	0.75	0.40

Table 3 presents the results obtained using only segment color histograms. To improve the received results the next step was to automatically detect the whole audience shots not only individual frames.

5.3 Selection of Audience Frames based on Green Color of Soccer Field and Shot Analysis

The observation was used that usually audience is presented by a stable camera, so, the similarity of the adjacent frames is significant. To find that two adjacent frames belong to the same shot two metrics were applied: Mean Squared Error (MSE) and Structural Similarity (SSIM) defined in (Wang et al., 2004; Ou et al.,2011). The following parameters were used:

- acceptable value of MSE – 5000,
- minimum number of frames in the audience shot – 40,
- minimum number of frames classified as audience frames in a shot – 75% of all frames in the shot,
- minimum number of frames classified as audience frames and not assumed as false detection – 10% of all frames in the shot.

As a consequence, the two consecutive frames were associated into the same video shot if the MSE value was not greater than 5000. Next a shot could not be shorter than 40 frames. If more than 75% of frames in the shot were selected as audience frames, the whole shot was assumed to be an audience shot. On the other hand if a few frames, less than 10% of frames in the shot, were selected as audience frames, the whole shot was assumed not to be an audience shot. Such a procedure improved the results of the selection of audience frames (Table 4). The precision increased from 0.59 to 0.69.

Table 4: Selection of audience frames using segment color histograms and shot analysis.

Maximum share [in %] of green color in a frame	Recall	Preci-sion	Fallout	Sensi-tivity	F-Measure
5	0.99	0.74	0.03	0.97	0.83
10	1.00	0.69	0.05	0.95	0.78
15	1.00	0.58	0.08	0.93	0.70
20	1.00	0.54	0.08	0.92	0.67
25	1.00	0.46	0.10	0.91	0.61
30	1.00	0.44	0.12	0.89	0.59
35	1.00	0.39	0.14	0.87	0.55
40	1.00	0.33	0.18	0.83	0.49

5.4 Selection of Audience Frames based on Face Detection

The last improvement tested in the experimental studies was based on the detection of faces (Mita et al., 2005). The assumption was that audience frames suitable for further content analysis are frames of medium views, where the fans are well distinguishable and their clothes, their different accessories like flags and banners as well as their gestures are also well seen. It will be ensured when we are able to detect faces of spectators in an analyzed frame. The question arises how many faces we can expect in one audience frame. The results were varied for three tested soccer sports videos. Tables 5–7 present individual results for all three videos.

Table 5: Selection of audience frames using face detection – Video 1.

Number of faces detected	Recall	Preci-sion	Fallout	Sensi-tivity	F-Measure
1	0.37	0.06	0.31	0.68	0.10
2	0.20	0.07	0.14	0.83	0.10
3	0.20	0.11	0.09	0.88	0.14
4	0.20	0.12	0.08	0.89	0.15
5	0.20	0.14	0.07	0.90	0.16
10	0.20	0.17	0.05	0.91	0.18
20	0.10	0.15	0.03	0.93	0.12
30	0.01	0.09	0.01	0.94	0.02

Table 6: Selection of audience frames using face detection – Video 2.

Number of faces detected	Recall	Preci-sion	Fallout	Sensi-tivity	F-Measure
1	0.35	0.14	0.22	0.74	0.20
2	0.32	0.29	0.08	0.86	0.30
3	0.30	0.55	0.03	0.91	0.39
4	0.09	0.65	0	0.91	0.16
5	0.08	0.76	0	0.91	0.14
10	0	0	0	0.91	0
20	0	0	0	0.91	0
30	0	0	0	0.91	0

Table 7: Selection of audience frames using face detection – Video 3.

Number of faces detected	Recall	Preci-sion	Fallout	Sensi-tivity	F-Measure
1	0.67	0.52	0.06	0.92	0.59
2	0.51	0.93	0	0.96	0.66
3	0.27	0.99	0	0.94	0.42
4	0.20	1.00	0	0.93	0.33
5	0.17	1.00	0	0.93	0.29
10	0.07	1.00	0	0.92	0.13
20	0	0	0	0.92	0
30	0	0	0	0.92	0

In the case of the Video 3 we can say about very promising results. If four or more faces were detected in a frame such a frame was correctly selected as an audience frame very useful for the further analysis of sports videos.

The observations of videos led to the conclusion that such a very specific characteristic of audience frames strongly depends on the way the audience is presented by a cameraman. However, for some videos such a procedure can be useful.

In these experiments the soccer videos were used. This approach can be easily applied also in the analysis of other sports categories where the playing fields have dominant colors. Green color is typical not only for soccer but also for example for American football. Tennis courts are not green, except may be for Wimbledon, but however they are characterized by one dominant color. So, this approach can be extended to the analysis of many other sports categories.

6 FINAL REMARKS

The content-based video analysis requires very sophisticated processing methods. Although many approaches have been proposed and still are being developed their efficiency is not satisfactory. The goal of the automatic analysis and indexing of sports videos is to recognize sports disciplines reported in sports news, detect highlights, generate summary, remove publicity segments or break parts, segment continuous sports videos, etc. These processes are using different strategies such as pattern frame comparison, line detection in playing fields, player detection, sports equipment detection, or detection of superimposed text, and others. In most approaches the audience shots are processed like other non-player shots, considered as not useful for video content analysis. In the research and experimental studies presented in this paper the audience shots in sports videos are considered as very informative parts helping to detect and recognize sports disciplines, nationality or club membership, as well as emotions of supporters.

Segment color histograms are usually applied to detect soccer playing fields and player shots. In this paper these histograms were used to detect audience frames and shots. The green color is not the dominant color for the audience of soccer games. Audience frames and shots present people in the stands at the stadium. But the dominant color as a unique criterion is not efficient in audience detection. This procedure was improved by analyzing not only single frames but sequences of frames belonging to one shot. Moreover, the application of a face detection method was tested to find the most suitable audience shots for content analysis of sports videos. The observations of sports videos suggest that the most informative audience shots are those recorded as medium views.

To analyze and to categorize the detected audience shots other approaches are envisaged to be applied such as crowd detection (Reisman et al., 2004; Wang et al., 2018) and people counting (Mousse et al., 2017). Because different groups of spectators can be present in the stands of the sports stadiums, for example two groups of fans of both teams of the soccer match, methods of group detection in crowded scenes could be also useful (Pandey et al., 2020).

The most significant advantage of audience shot detection is the opportunity to analyze the behavior of spectators, their club clothes, flags and banners, and even their gestures and expressions of joy, and then categorize and better annotate sports videos.

REFERENCES

Assfalg, J., Bertini, M., Colombo, C., and Del Bimbo A. (2002). Semantic annotation of sports videos, *IEEE Multimedia*, 9 (2): 52–60.

Bertini M., Del Bimbo, A., and Nunziati, W. (2005). Automatic annotation of sport video content, in: Lazo M. and Sanfeliu A. (Eds.), *CIARP*, LNCS, vol. 3773, pp. 1066–1078.

Choroś K. (2012). Video structure analysis for content-based indexing and categorisation of TV sports news. *International Journal of Intelligent Information and Database Systems*, 6(5): 451–465.

Choroś K. (2016). Weighted indexing of TV sports news videos. *Multimedia Tools and Applications*, 75(24): 16923–16942.

Choroś, K. (2010). Video structure analysis and content-based indexing in the Automatic Video Indexer AVI, in: *Advances in Multimedia and Network Information System Technologies, Advances in Intelligent and Soft Computing*, Springer, AISC, vol. 80, pp. 79–90.

Choroś, K. (2012). Detection of tennis court lines for sport video categorization, in: *Computational Collective Intelligence. Technologies and Applications*, Springer, LNCS, vol. 7654, pp. 304–314.

Daudpota, S.M., Muhammad, A., and Baber, J. (2019). Video genre identification using clustering-based shot detection algorithm. *Signal, Image and Video Processing*, 13(7): 1413–1420.

Ekin, A., Tekalp, A.M., and Mehrotra, R. (2003). Automatic soccer video analysis and summarization. *IEEE Transactions on Image Processing*, 12(7): 796–807.

Fang, T., and Ping, S. (2013). Attractive events detection in soccer videos based on identification of shots," in: *Proceedings of the 3rd International Conference on Multimedia Technology ICMT-13*, Atlantis Press, pp. 814–822.

Jiang, H., and Zhang, M. (2011). Tennis video shot classification based on support vector machine, in: *Proceedings of the IEEE International Conference on Computer Science and Automation Engineering*, IEEE, vol. 2, pp. 757–761.

Kuo, C.M., Chang, W.H., Fang, M.Y., and Lin, C.H. (2011) A template-based baseball video scene classification using efficient playfield segmentation. *Multimedia Tools and Applications*, 55(3): 399–422.

Li, Y., and Dorai, C. (2005). Video frame identification for learning media content understanding, in: *Proceedings of the IEEE International Conference on Multimedia and Expo*, IEEE, pp. 1488–1491.

Mita, T., Kaneko, T., and Hori, O. (2005). Joint Haar-like features for face detection, in: *Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05)*, IEEE, vol. 2, pp. 1619–1626.

Mousse, M.A., Motamed, C., and Ezin, E.C. (2017). People counting via multiple views using a fast information fusion approach. *Multimedia Tools and Applications*, 76(5): 6801-6819.

Ou, T.S., Huang, Y.H., and Chen, H.H. (2011). SSIM-based perceptual rate control for video coding, *IEEE Transactions on Circuits and Systems for Video Technology*, 21(5): 682–691.

Pandey, M., Singhal, S., and Tripathi, V. (2020). An efficient vision-based group detection framework in crowded scene. in: *Frontiers in Intelligent Computing: Theory and Applications,* Springer, AISC, vol. 1014, pp. 201–209.

Reisman, P., Mano, O., Avidan, S., and Shashua, A. (2004). Crowd detection in video sequences, in: *Proceedings of the IEEE Intelligent Vehicles Symposium*, IEEE, pp. 66–71.

Shih H.-C. (2017). A survey of content-aware video analysis for sports. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(5): 1212–1231.

Thomas, G., Gade, R., Moeslund, T.B., Carr, P., and Hilton, A. (2017). Computer vision for sports: Current applications and research topics. *Computer Vision and Image Understanding*, 159: 3–18.

Wang, Z., Bovik, A.C., Sheikh, H.R., and Simoncelli E.P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612.

Wang, Z., Cheng, C., and Wang, X. (2018). A fast crowd segmentation method, in: *Proceedings of the International Conference on Audio, Language and Image Processing ICALIP*, IEEE, pp. 242–245.

Yakut, M., and Kehtarnavaz, N. (2016). Ice-hockey puck detection and tracking for video highlighting. *Signal, Image and Video Processing*, 10(3): 527–533.

Zhang, Y.Z., Dong, Q., Wang, J.Y., and Dai, Y.W. (2011). Court view shots detection based on Hough transform and SVM, in: *Proceedings of the International Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping*, IEEE, pp. 1–4.