ROBUST DOMINANT COLOR REGION DETECTION AND COLOR-BASED APPLICATIONS FOR SPORTS VIDEO

Ahmet Ekin[1] *and A. Murat Tekalp*[1,2]

[1]Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY, 14627
[2]College of Engineering, Koc University, Istanbul, Turkey
{ekin,tekalp}@ece.rochester.edu

ABSTRACT

This paper proposes a novel automatic dominant color region detection algorithm that is robust to temporal variations in the dominant color due to field, weather, and lighting conditions throughout a sports video. The algorithm automatically learns the dominant color statistics of the field independent of the sports type, and updates color statistics throughout a sporting event by using two color spaces, a control space and a primary space. The robustness of the algorithm results from adaptation of the statistics of the dominant color in the primary space with drift protection using the control space, and fusion of the information from two spaces. We also propose novel and generic color-based algorithms for referee, player-of-interest, and play-break event detection in sports video. The efficiency of the proposed algorithms is demonstrated over a dataset of various sports video, including basketball, football, golf, and soccer video.

1. INTRODUCTION

Color is an essential feature in sports video. Playing field in most sports can be characterized by a single dominant color, and the players and referee wear distiguishable colored uniforms. In this paper, we focus on robust detection of field region since accurate segmentation of field region yields improvements in the performance of higher level algorithms, including player/referee and play-break event detection.

Since the field color demonstrates variations during a sports broadcast due to changes in *imaging conditions,* such as viewing direction, and *environmental conditions,* such as shadows and sunset, it is important to adapt field color statistics to these variations. Popular color spaces for field region extraction have been hue-saturation variants, such as HSI and HSV [1, 2] and RGB [3]. However, the adaptation of color statistics to changing imaging conditions is neglected except for [1], which updates green color for football video. In contrast, we propose a novel generic dominant color region detection algorithm with a primary space and a control space for sports video. The proposed algorithm automatically detects the dominant color of the field independent of the sports type, and updates the statistics of the dominant color to the variations in imaging conditions throughout a game.

In the next section, we explain the proposed dominant color region detection algorithm. After that, in Sec. 3, we present novel algorithms for detection of referee, players-of-interest, and play-break events. The effectiveness of the proposed algorithms are demonstrated over basketball, football, golf, and soccer video in Sec. 4.

2. DOMINANT COLOR REGION DETECTION

Playing field can be described by *one distinct dominant color.* This dominant color, however, demonstrate variations from one sport to another, from one stadium to another, and sometimes within one stadium during a sporting event. In a specific sporting event, variations in the dominant field color are caused by changes in environmental factors, such as sunset, shadows, rain, and snow, spatio-temporal variations in illumination intensity due to irregularly positioned and/or imperfect stadium lights, and spatial variations in field properties, such as grass color variations and field deformations. We propose a new algorithm that learns the statistics of the dominant color, adapts these statistics to changing imaging conditions for robust performance, and detects dominant color region independent of the sports type.

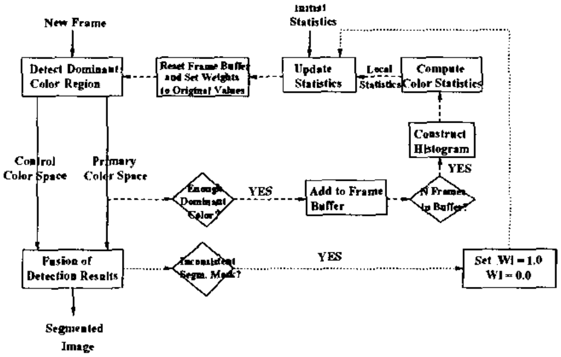


Fig. 1. The flowchart of the proposed dominant color region detection algorithm

2.1. Proposed Algorithm

The proposed generic algorithm automatically extracts dominant color statistics and adapts them to the variations in imaging conditions that are typical for sports video. We observed that none of the

This work is supported in part by National Science Foundation under grant number IIS-9820731 and in part by NYSTAR grant to the Center for Electronic Imaging and Science, Rochester, NY.

0-7803-7750-8/03/$17.00 ©2003 IEEE

color spaces, if they are used alone, performs well under all variations in sports video. Furthermore, when intensity information is used, the resulting segmentation map is cleaner, but the statistics need adaptation, which may cause drifts from the actual dominant color statistics in the long term. Therefore, we propose to use two color spaces, which should be *complementary* and *essentially different* color spaces for robustness. One of these spaces, *control space,* is employed to control the other space, called primary space, which is adapted to local variations; hence, color spaces that include intensity information, such as HSI or La^*b^*, may be employed as the primary space. Control space serves for two purposes: 1) It prevents the local statistics to dominate and eventually drift from the true statistics in the primary space, and 2) the fusion of segmentation masks in primary and control color spaces increases the pixelwise accuracy in dominant color region detection.

In Fig. 1, the flowchart of the proposed algorithm is given. At start-up, the system computes initial statistics and the values of several parameters for each color space from the frames in the training set (this step has not been included to Fig. 1). After the initialization of parameters, dominant color region for each new frame is detected in both control and primary color spaces. Segmentation results in these spaces are used by the fusion algorithm, explained in Sec. 2.3, to obtain pixelwise more accurate final segmentation mask. The rest of the blocks in the flowchart are utilized for adaptation of primary color space statistics by two feedback loops. The inner feedback loop, connected with the dashed lines, computes local statistics in primary color space and captures local variations, whereas the other feedback loop, connected with the dotted lines, becomes active when segmentation results conflict with each other, which indicates drifting of local statistics from true statistics in primary color space. The activation of this outer feedback loop resets primary color statistics to their initial values.

2.2. Distance Metrics

We employ two distance metrics, Euclidean metric (L_2 norm) for all spaces except for HSI, for which *robust* cylindrical metric, defined in Eqs. 1-3, is used [4].

$$d_I(j,k) = |I_j - I_k| \tag{1}$$

$$d_{chroma}(j,k) = \sqrt{(S_j)^2 + (S_k)^2 - 2S_jS_kcos(\theta)} \tag{2}$$

$$d_{cylindrical}(j,k) = \sqrt{(d_I(j,k))^2 + (d_{chroma}(j,k))^2} \tag{3}$$

In Eqs. 1-3, S and I refer to saturation and intensity, respectively, j and k are the j^{th} and k^{th} pixel, and θ is the minimum absolute difference between the two hue values, i.e., θ is limited to $[0,\pi]$. For achromatic dominant color regions and pixels, the distance $d_{cylindrical}$ is taken to be equal to intensity distance d_I.

2.3. Dominant Color Region Detection and Fusion of Results

Dominant field color is described by mean values of each color component, such as r_M and g_M for rg, and H_M, S_M, and I_M for HSI. Assuming these parameters are computed at the start-up and updated, the current frame is segmented by finding the color distance of each pixel to the mean values in both spaces. If the color distance of a pixel to the dominant color is less than T^C_{color} in control space (T^P_{color} in primary color space), then the pixel is assigned as a field pixel in the corresponding color space. The parameters T^C_{color} and T^P_{color} are the maximum allowable distance values in control and primary color spaces, respectively.

Dominant color region detection results are fed into the fusion algorithm. Denoting these input masks as M_C and M_P, segmentation results in control and primary spaces, respectively, and the output mask from the fusion algorithm as M_F, we define G_C and G_P as dominant colored pixel ratios in M_C and M_P, respectively. The proposed fusion algorithm determines the consistency of the results by verifying the absolute difference of G_C and G_P to be small. When there is inconsistency, the algorithm assumes that the result in primary color space is unreliable, discards M_P, and uses control space dominant color region mask, M_C, as the final result, M_F. In this case, the second feedback loop, linked with dotted lines in Fig. 1, are activated to reset the primary color space statistics to their initial values. When the results are consistent, the fusion algorithm selects the mask image with the smallest number of dominant colored pixels if either G_C or G_P is small. This is because the conditions are indicative of an out-of-field or a close-up view and selecting the mask with lower dominant colored pixel ratio increases the segmentation accuracy. In the opposite case, i.e., when G_C or G_P are large, the fusion algorithm selects the mask image with the largest number of dominant colored pixels. For other cases, the result of the primary color space is used.

2.4. Computation and Adaptation of Color Statistics

Dominant color statistics are computed for two purposes: 1) to initialize the statistics in control and primary color spaces from the training set, and 2) to adapt the statistics in the primary color space to variations in imaging conditions (*Compute Color Statistics* block in Fig. 1). Since dominant color is described by mean color, the algorithm estimates the mean values of each color component from average color histograms of either the frames in the training set (initialization) or the frames in the buffer (adaptation). The mean values are computed as the mean of the histogram intervals about peak indices for each one-dimensional average histogram. The interval definition is similar to alpha-trimmed histograms and the details are in [5]. During initialization, the parameters, T^C_{color} and T^P_{color}, defined in Sec. 2.3, are also computed if *a rough boundary of the field region* or *the dominant colored pixel ratio for the frames in the training set* is given. Then, T^C_{color} and T^P_{color} are adjusted to classify all pixels, other than the outliers, in the bounding box or the given percentage of pixels in the whole frame in the dominant color region class.

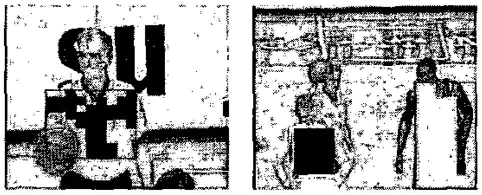
Primary color statistics are updated by either the inner or both the inner and the outer feedback loops. The inner loop (dashed lines in Fig. 1) computes local statistics from the histograms of the last N selected frames that have enough number of dominant colored pixels. When N frames are accumulated in the buffer, the average histogram and the mean values of color components from the average histogram are computed as explained above. The statistics in primary color space are updated as the sum of weighted local and initial statistics as shown in Eq. 4, where w_l and w_i denote the weights for local color mean vector, $\mathbf{M_{local}}$, and initial color mean vector, $\mathbf{M_{initial}}$, respectively. As shown in Fig. 1, w_l becomes zero when local statistics become unreliable, i.e., when the outer feedback loop is activated.

$$\mathbf{M_{curr}} = w_l * \mathbf{M_{local}} + w_i * \mathbf{M_{initial}} \tag{4}$$

3. APPLICATIONS

3.1. Detection of Referee and Players-of-Interest

Color features of referee and player uniforms are used for detection and localization of these objects because the official rules of most team sports require dominant colors of team and referee uniforms to be distinguishable from each other. In the proposed algorithm, we describe the shirt colors of each team and referee by color histograms, H_{TeamA}, H_{TeamB}, and H_{ref}, respectively. Then, the algorithm searches each medium or close-up frame for the most salient object, i.e., the object of interest, from each class. Since the locations of these objects in a frame are not known a priori and it is unlikely that a frame consists only of these objects, each frame is divided into overlapping blocks of size MxN, where the overlapping is achieved by moving the center of the first block by $m * M/2$ and $n * N/2$ for integers m and n to span the whole frame. The similarity of color histogram of each block to each class, i.e., two teams and referee, is found by histogram intersection and if the maximum similarity score among the three classes is greater than a parameter, T_{sim}, that block is assigned to the class of the histogram that gives the maximum similarity value. In Fig. 2, detected referee and team blocks are shown for example images.



(a) (b)

Fig. 2. a) Referee uniform colored blocks and the bounding box for the detected referee region, and b) detected players of interest from each team

To detect referee and a player of interest from each team, the largest number of connected blocks from each class is found. The minimum bounding rectangle (MBR) around these blocks is defined, shown in Fig. 2, for each class and verified against *the similarity of MBR_{obj} histogram to the corresponding class, the ratio of the area of the MBR_{obj} to the frame area,* and *MBR_{obj} aspect ratio (width/height).*

3.2. Play-Break Event Detection

Sports video is composed of play and break events, which may be frequent and may take a considerable amount of broadcasting time in some sports, such as football and baseball. Therefore, detection of play and break events makes it possible to generate concise lossless summaries consisting of all play events in a specific game. Sports video is composed of long, medium, and close-up shots [5]. Among these shots, long shots usually correspond to play events, while close-up shots indicate breaks in the game. However, broadcasters also use long shots during a break. Therefore, in addition to *shot-type*, we also use *shot length* to locate play events. For that purpose, minimum long shot length, L_{min}, that refers to a play event is computed from the training videos. Then, as shown in Fig. 3, play events in long shots are localized by comparing the length of each long shot against L_{min}, and labeling each long shot having length longer than L_{min} as play.

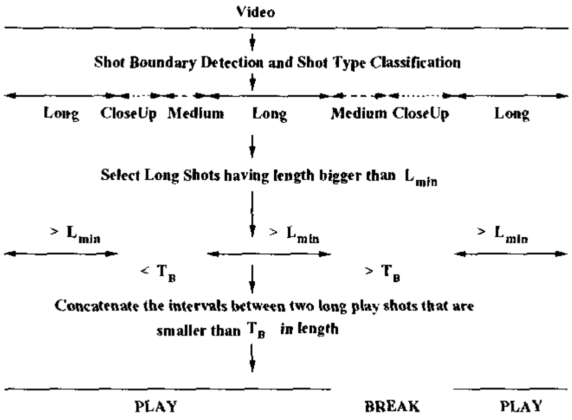

Video
Shot Boundary Detection and Shot Type Classification
Long
CloseUp
Medium
Long
Medium
CloseUp
Long
Select Long Shots having length bigger than L_{min}
$> L_{min}$
$> L_{min}$
$> L_{min}$
$< T_B$
$> T_B$
Concatenate the intervals between two long play shots that are smaller than T_B in length
PLAY
BREAK
PLAY

Fig. 3. A sample play-break event detection

Play events are usually captured as long shots, but medium shots or player close-ups may be intermittently used to highlight certain objects for a short time. In order to detect such play events, we use another parameter, T_B, defined as the maximum allowable time between two long shot play events. If the interval between two long shot play events is shorter than T_B sec, the corresponding segment is labeled as a play. Otherwise, it is described as a break. The proposed play-break event detection algorithm is *generic* in that it uses domain information through only L_{min} and T_B that can be passed as parameters to the same algorithm for different types of sports. In the proposed algorithm, the value of T_B determines the actual compression rate for the generated summary; as T_B increases, the summary length increases since some break events are also included to the summaries. In Sec. 4, we elaborate on its effect in the compression rate and in the play-break event detection accuracy.

4. RESULTS

4.1. Dominant Color Region Detection

In Table 1, the performance of dominant color region detection algorithm is tabulated for each color space and sports type over a large dataset of sports video. We define the error measure, *segmentation inaccuracy,* as the ratio of the number of misclassified pixels to the total number of pixels. Table 1 shows that the proposed two-space algorithm improves the results obtained when primary color space is used alone, which indicates that the use of control space prevents drifts from true color means and/or the fusion algorithm improves the segmentation accuracy. The results also demonstrate that the best combination is $rg - HSI$ pair, where the former is selected as the control and the latter as primary color spaces. This may be because 1) hue, which is used for adaptation, is invariant to highlights, 2) the cylindrical distance metric for HSI usually provides superior results over Minkovski distance metrics [4], and 3) rg and HSI tend to be complementary in sports video, when one fails, the other usually provides accurate detection. Single color spaces are outperformed by rg and HSI combination for all cases, except for the golf sequence, where $CrCb$ space generated

Color Space(s)	Segm. Inaccuracy ([0.0,1.0])			
Control + Primary	Basketball	Football	Golf	Soccer
$CrCb$	0.18	0.06	0.02	0.12
rg	0.12	0.07	0.09	0.08
HSI	0.17	0.03	0.07	0.11
La^*b^*	0.17	0.06	0.06	0.12
$rg + HSI$	0.04	0.01	0.03	0.01
$CrCb + HSI$	0.16	0.03	0.02	0.08
$rg + La^*b^*$	0.09	0.04	0.06	0.04
$CrCb + La^*b^*$	0.11	0.04	0.04	0.11

Table 1. The effect of using different color spaces in segmentation

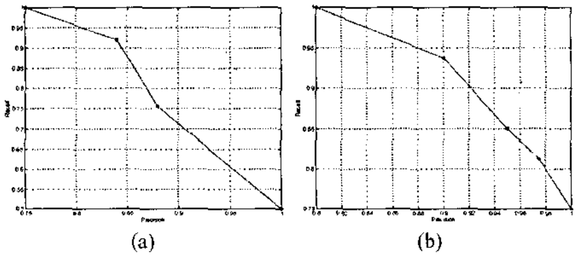
T_B	SpainB	NCAAB	KoreaB
(sec)	D, FA	D, FA	D, FA
5	0.92, 0.0	0.92, 0.08	0.81, 0.0
10	0.92, 0.0	0.95, 0.08	0.92, 0.0
20	0.98, 0.23	1.0, 0.39	1.0, 0.18
30	1.0, 0.25	1.0, 0.39	1.0, 0.18

Table 2. Performance of the proposed play-break event detection algorithm for basketball video (D: play event detection rate, FA: play event false alarm rate)

slightly better result. More importantly, the algorithm with rg and HSI choices is consistent throughout the whole data set, achieving more than 95% segmentation accuracy for all video clips. Finally, fusion of information from two color spaces improves the segmentation accuracy by 3% on the average.

4.2. Referee and Player-of-Interest Detection

In Fig. 4, recall-precision values are plotted by changing T_{sim}, defined in Sec. 3.1. Fig. 4 (a) shows that the algorithm achieves more than 80% precision and 90% recall rates for referee detection. The same algorithm determines one player-of-interest for each team, and as shown in Fig. 4 (b), more than 90% precision and recall rates can be achieved at the same time. Since the color of referee uniform is in general less striking than the colors of team uniforms, the precision rate for referee detection at the same recall value is lower.



(a) (b)

Fig. 4. Precision-Recall plots for detection of a) referee, b) players of interest

4.3. Play-Break Event Detection

Table 2 shows detection (D) and false alarm rate (FA) over three basketball sequences for various T_B values. In the experiment, the ground truth play-break event labels were assigned by a human operator. These ground truth play-break segments achieve 30.9, 22.8, and 18.9% compression rates for *SpainB*, *NCAAB*, and *KoreaB* sequences, respectively. Table 2 shows that the optimal T_B value is dependent on a particular sequence and a particular broadcaster. For example, in order to capture all plays, T_B should be 30 sec for *SpainB* sequence, 20 sec for the *NCAAB* and *KoreaB* sequences. In spite of these variations, it is still possible to conclude that play event summaries generated for T_B values at the lower half of 10-20 sec interval consist of 92-95% of all play events with a very low false alarm rate, while the upper half of the same interval results in the summaries of almost all, 98-100%, play events with some false positives.

We also used the same algorithm for football video by setting T_B to 5 sec because football play events are shorter. The proposed algorithm detected all play events (D = 1.0) in the football clip with a low false alarm rate, FA = 0.02, achieving a compression rate of 68%, which, due to the characteristics of football games, is more than twice the maximum compression rate for basketball clips. This also shows that play-break event detection is very useful and has immediate commercial implications for football video summarization while it is more applicable as a pre-processing tool for basketball video.

5. CONCLUSION

We proposed a new robust and generic dominant color region detection algorithm with a primary space and a control space for sports video. We showed the robustness of the algorithm to variations in imaging conditions in various sports. Since the proposed algorithm is the first preprocessing step in high-level sports video analysis, its robustness is essential for an accurate and reliable semantic analysis. We also presented novel algorithms for referee and player-of-interest detection as well as play-break event detection.

6. REFERENCES

[1] B. Li and M. I. Sezan, "Event detection and summarization in American football broadcast video," in *Proc. SPIE*, Jan. 2002.

[2] J. Assfalg, M. Bertini, C. Colombo, and A. del Bimbo, "Semantic annotation of sports videos," *IEEE Multimedia*, 9(2):52-60, Apr.-June 2002.

[3] N. Babaguchi, Y. Kawai, and T. Kitashi, "Event based indexing of broadcasted sports video by intermodal collaboration," *IEEE Trans. on Multimedia*, 4(1):68-75, March 2002.

[4] K. N. Plataniotis and A. N. Venetsanopoulos, *Color image processing and applications*, Springer-Verlag, Berlin, Germany, pp. 25-32 and 260-275, 2000.

[5] A. Ekin and A.M. Tekalp, "Robust dominant color region detection with applications to sports video," submitted to *CVIU*, also at *http://www.ece.rochester.edu/~ekin*