Color Video Segmentation using Fuzzy C-Mean Clustering with spatial information

M. ARFAN JAFFAR, BILAL AHMED, NAWAZISH NAVEED, AYYAZ HUSSAIN, and ANWAR M. MIRZA
Department of Computer Science
National University of Computer and Emerging Sciences
A.K. Brohi Road H-11/4, Islamabad,
PAKISTAN
arfan.jaffar@nu.edu.pk, balodhi@hotmail.com , nawazishnaveed@yahoo.com, ayyaz.hussain@nu.edu.pk, anwar.m.mirza@nu.edu.pk

Abstract: - Video segmentation can be considered as a clustering process that classifies one video succession into several objects. Spatial information enhances the quality of clustering which is not utilized in the conventional FCM. Normally fuzzy c-mean (FCM) algorithm is not used for color video segmentation and it is not robust against noise. In this paper, we presented a modified version of fuzzy c-means (FCM) algorithm that incorporates spatial information into the membership function for clustering of color videos. We used HSV model for decomposition of color video and then FCM is applied separately on each component of HSV model. For optimal clustering, grayscale image is used. Additionally, spatial information is incorporated in each model separately. The spatial function is the summation of the membership function in the neighborhood of each pixel under consideration. The advantages of this new method are: (a) it yields regions more homogeneous than those of other methods for color videos; (b) it reduces the spurious blobs; and (c) it removes noisy spots. It is less sensitive to noise as compared with other techniques. This technique is a powerful method for noisy color video segmentation and works for both single and multiple-feature data with spatial information.

Key-words: color video segmentation, spatial fuzzy c-mean, *and cluster validity, Frame change detection*

1. INTRODUCTION

Video segmentation has been a significant and challenging problem for many video applications [1,2]. It is a key step in image sequence analysis that is extensively used for determining motion features of scene objects, as well as for coding purposes to reduce storage requirements. It can be considered as a clustering process that classifies one video succession into several objects [3]. In a video, each object can be considered as one pattern represented by spatial and/or temporal features. There are different and ambiguous definitions of video objects e.g., color, texture, or motion. Most automatic object-based segmentation methods use these criteria to define video objects and we have used color for automatic segmentation of object in a video [12,13].

A video stream is a temporally emerging medium, where content changes occur due to camera or object motion, cuts, and special effects. Temporal video segmentation, which constitutes the first step in content-based video analysis, refers to breaking the input video into temporal segments with uniform content. Video segmentation is different from segmentation of single image. While several correct solution may exist for segmenting a single image. Video segmentation is fundamental step towards structured video representation, which supports the interpretability and manipulability of visual data

Fuzzy c-means (FCM) clustering [4,5,6,14] is an unsupervised method that has been successfully applied to feature analysis, clustering, and classifier designs in fields such as astronomy, geology, medical imaging, target recognition, and image segmentation. An image can be represented in various feature spaces, and the FCM algorithm classifies the image by grouping similar data points in the feature space into clusters. This clustering is achieved by iteratively minimizing a cost function that is dependent on the distance of the pixels to the cluster centers in the feature domain.

One of the important characteristics of an image is that neighboring pixels are highly correlated. In other words, these neighboring pixels possess similar feature values, and the probability that they belong to the same cluster is great. This spatial relationship is important in clustering, but it is not utilized in a standard FCM algorithm.

To exploit the spatial information, a spatial function is used, where neighborhood represents a square window centered on pixel x_j in the spatial domain. A 3X3 window was used throughout this work. Just like the membership function, the spatial function h_{ij} represents the probability that pixel x_j belongs to i^{th} cluster. The spatial function of a pixel for a cluster is huge if the majority of its neighbor belongs to the similar clusters. Hence, it is robust against noise.

The paper is organized as follows. Section 2 has related work information. Section 3 describes the proposed method. Section 4 contains implementation of our novel technique. Section 5 describes the results. Section 6 describes the conclusion and future work. Section 7 composed of the references.

2. RELATED WORK

Layering of image segmentation was introduced by Darrell and Pentland [7], Wang and Adelson [8]. Darrel and Pentland [7] estimated the number of layers and the pixel assignments to each layer through robust estimation. Examples are shown with range images and optical flow. The seminal paper on segmenting video into layers is marked as Wang and Adelson [8]. Affine model is adjusted to blocks of optical flow which follows K means clustering of the affine parameters. In this stage splitting and merging of layers is done and thus it is not fixed. The shape of regions is not limited to aggregate of blocks after the first iteration, but is taken to a pixel level within the blocks. Convincing results are prompted; despite of the edges of segments not being very accurate and the reason might be the errors in the computation of optical flow at occlusion boundaries. Bergen et. al. [9] show a method for motion segmentation in which the global parametric motion of the entire frame is computed first and then finding the segments that do not adjust well in the global motion model. In this loop Irani and Peleg [10] add temporal integration. To register the current image with the previous one, weighted aggregate of a number of frames is used. Thus the object becomes sharply into focus and everything else blurs out which improves the stability of the solution. Thompson [11] depicts the earliest work on combining multiple features for segmentation. The image is segmented based on intensity and motion, by finding 4-connected regions which are then merged together using a variety of heuristics.

3. METHODOLOGY

Out proposed method is based on different modules.

3.1 Preprocessing

First of all as a preprocessing step, video data is converted into frames and stored for further accesses in different functions.

There are different and ambiguous definitions of video objects e.g., color, texture, or motion. So each frame can have a number of clusters. We do not know exactly how many clusters in each frame. So a mechanism is needed to find out optimal number of clusters. We have used FCM and different cluster validity functions to calculate optimal number of clusters of all frames so that during segmentation these optimal numbers of clusters can pass to FCM. Moreover we have also used these clusters as frame change detection.

3.2 Frame Change Detection

We have used two different measures to find out frame change detection. First of all, we have used optimal number of clusters of two frames to check difference of these two frames. If both frames have different number of optimal clusters then there is a need to segment both frames. Otherwise there is possibility that both frames have no difference. In case that both frames have same number of clusters then we find out again color histogram of both frames by using equation that return the histogram difference eq. (1) of both frames. If that difference is smaller than a specific threshold then both frames have same scene and there is no change in both frames. In this case we will segment the first frame and skip next frame so that efficient segmentation can be performed.

$$HD_{(k,k+1)} = \sum_{i=0}^{G-1} |H_{k+1}(i) - H_k(i)| \tag{1}$$

3.3 Frames Segmentation

Color frame is taken as input to this block and converted to HSV color model image. In parallel 3 layers of HSV model are extracted from color image. Each layer of the image is passed through a two-pass process for clustering.

a. *FCM Clustering*

Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to be in the right position to two or more clusters. This method (developed by Dunn in 1973 and improved by Bezdek in 1981) is frequently used in pattern recognition. FCM starts with an initial guess for the cluster centers, which are proposed to mark the mean gns every data point a membership rank for every cluster. By iteratively updating the cluster centers and the membership grades for each data point, FCM iteratively moves the cluster centers to the correct place within a dataset. This iteration is based on minimizing an objective function that symbolizes the distance from any given data point to a cluster center weighted by that data point's membership rank.

The FCM algorithm assigns pixels to each group by using fuzzy memberships. Let XZ(x1, x2,., xN) indicates an image with N pixels to be partitioned into c clusters, where xi represents multispectral (features) data. The algorithm is an iterative optimization that minimizes the cost function defined as follows:

$$J = \sum_{j=1}^{N} \sum_{i=1}^{C} U_{ij}^{m} \left\| x_j - V_i \right\|^2$$

Where uij represents the membership of pixel xj in the ith cluster, vi is the ith cluster center, ||..|| is a norm metric, and m is a constant. The parameter m controls the fuzziness of the resulting partition, and m=2 is used in this study. The cost function is minimized when pixels close to the centroid of their clusters and are assigned high membership values, and low membership values are assigned to pixels with data far from the centroid. The membership function represents the probability that a pixel belongs to a specific cluster. In the FCM algorithm, the probability is dependent solely on the distance between the pixel and each individual cluster center in the feature domain. The membership functions and cluster centers are updated by the following:

$$U_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{x_j - v_i}{x_j - v_k} \right)^{2/(m-1)}}$$

and

$$v_i = \frac{\sum_{j=1}^{N} U_{ij}^{m} x_j}{\sum_{j=1}^{N} U_{ij}^{m}}$$

Starting with an initial guess for each cluster center, the FCM converges to a solution for vi representing the local minimum or a saddle point of the cost function. Convergence can be detected by comparing the changes in the membership function or the cluster center at two successive iteration steps.

b. *sFCM Method*

One of the significant uniqueness of an image is that neighboring pixels are extremely correlated. In other terms these neighboring pixels hold similar feature values, and the probability that they belong to the same cluster is great. This spatial relationship is important in clustering, but it is not utilized in a standard FCM algorithm. To develop the spatial information, a spatial function is defined as

$$h_{ij} = \sum_{k \in NB(x_j)} U_{ik} \quad (4)$$

Where NB(xj) stands for a square window centered on pixel xj in the spatial domain. A 3*3 window was used throughout this effort. Just like the membership function, the spatial function hij stands for the probability that pixel xj belongs to ith cluster. The spatial function of a pixel for a cluster is large if the bulk of its neighborhood belongs to the same clusters. The spatial function is included into membership function as follows:

$$U'_{ij} = \frac{U_{ij}^{p} h_{ij}^{q}}{\sum_{k=1}^{C} U_{kj}^{p} h_{kj}^{q}} \quad (5)$$

where p and q are parameters to control the relative importance of both functions. In a homogenous region, the spatial function simply fortifies the original membership, and the clustering result remains unchanged. However, for a noisy pixel, this formula reduces the weighting of a noisy cluster by the labels of its neighboring pixels. As a result, misclassified pixels from noisy regions or spurious blobs can easily be corrected. We separately applied this sFCM on each component of HSV model. The spatial FCM with parameter p and q is denoted sFCMp,q. Note that sFCM1,0 is identical to the conventional FCM. The clustering is a two-pass process, at each iteration for each component of the HSV model. The first pass is the same as that in standard FCM to calculate the membership function in the spectral domain. In the second pass, the membership information of each pixel is mapped to the spatial domain, and the spatial function is computed from that. The FCM iteration proceeds with the new membership that is incorporated with the spatial function. The iteration is stopped when the maximum difference between two cluster centers at two successive iterations is less than a threshold (0.02). After the convergence, defuzzification is applied to assign each pixel to a specific cluster for which the membership is maximal. [4]

In first pass simple FCM algorithm is applied on each layer for calculating the fuzzy membership. And in the second pass spatial information is incorporated in the calculated fuzzy membership of each layer. A 3x3 window is used throughout this work for incorporating spatial information. After sFCM layers are defuzzified to get the crisp set of values. HSV layers are then combined and converted to RGB model for generating the clustered image.

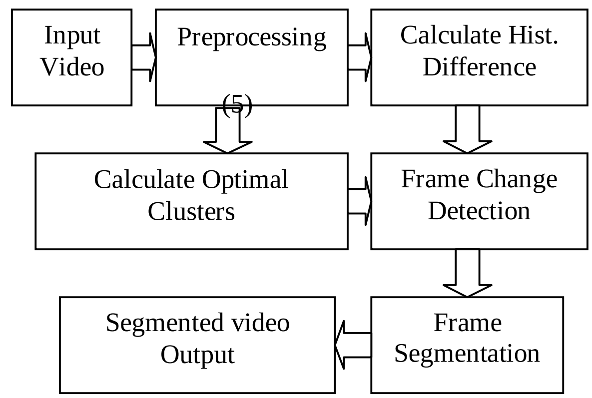
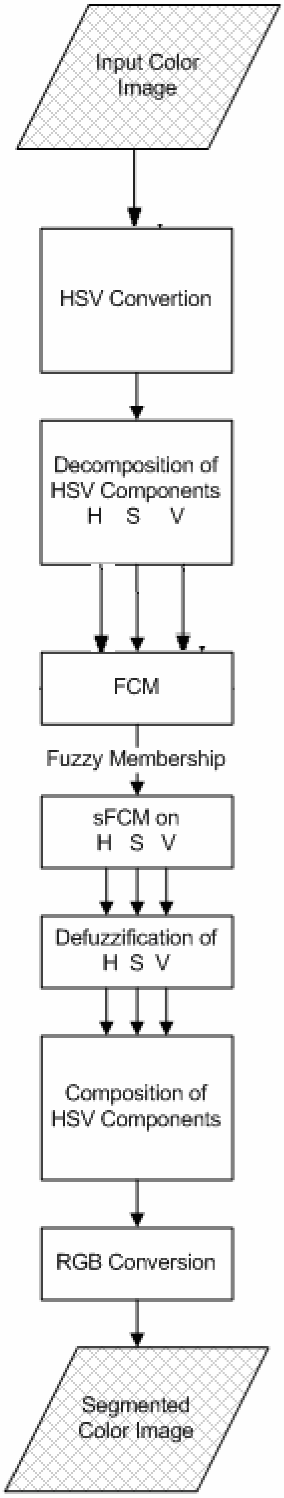


Figure 1: Block diagram of Video Segmentation

4. IMPLEMENTATION

This algorithm is implemented in Matlab®. The frame segmentation process model is explained in a flowchart diagram 2. First of all, preprocessing has been done on input video. Then frames can be converted into color images. We calculate the optimal clusters of each frame and also calculate the histogram difference. If optimal clusters of two consecutive frames have different, then we apply directly segmentation on both frames otherwise we check histogram difference. If histogram difference is less then some threshold then we skip that frame.

Color image is taken as input to the system and converted to gray scale image for extracting optimal



Flowchart Diagram 2: General flowchart

number of clusters by applying FCM algorithm. In parallel 3 layers of HSV model are extracted from color image. Each layer of the image is passed through a two-pass process for clustering. In first pass simple FCM algorithm is applied on each layer for calculating the fuzzy membership. And in the second pass spatial information is incorporated in the calculated fuzzy membership of each layer. A 3x3 window is used throughout this work for incorporating spatial information. After sFCM layers are defuzzified to get the crisp set of values. HSV layers are then combined and converted to RGB model for generating the clustered image.

5. RESULTS

Although we tested our algorithm over a large number of videos with varying range of complexity, here we show the experimental results for two videos only. Visual results are shown in figure [3] and figure [4].

We have implemented the proposed system by using the MATLAB environment. In this work, we have studied the performance of different segmentation techniques that are used in color video segmentation. These methods give good results on test databases of reasonable size. The number of clusters varies from frame to frame so there is need a mechanism to find out optimal number of clusters. These factors make the selection of gray-level segmentation threshold difficult, as different thresholds are likely required for different subjects.

Figure 4(b) gives the result of our proposed technique on the test image. It is evident through observation that the proposed system produces much smoother results than the schemes that have been used earlier. It can be proved that nearly all previous techniques don't work for the images when there is overlapping of intensities in lung parenchyma and surrounding chest wall. However, out proposed technique has shown promising results on the different test cases as shown in figure. Results of our proposed method that are shown in fig 4(b) demonstrate significant improvement.

In our experiments, we have used two test sequences that present different challenges to the video segmentation problem. The flower sequence contains objects moving only due to camera motion, and thus their motion is smooth and consistent. Portions of the image in this sequence consist of small texture, with a number of colors in it. Thus, color segmentation does not yield satisfactory results for object based compression. Motion segmentation of this sequence yields errors too, at the boundaries of the tree. The second test sequence is the table-tennis sequence. Its characteristics are almost entirely opposite to the flower sequence. There is almost zero background motion in the first part of the sequence. The texture is largely uniform and smooth, and therefore color segmentation alone does a decent job of extracting the objects of interest. Motion segmentation, on this sequence, suffers from temporal consistency problems, because the player repeatedly moves and stops. Therefore the player is very visible in the magnitude of optical flow for a few frames and then disappears for a few frames. Results are presented sequences using the method described in this paper. It can be seen that the method works well to correct both the errors mentioned above. It corrected the occlusion boundaries errors in flower sequence, so that the tree segment has accurate boundaries. It also keeps the player segmented out even when he stops and has the same optical flow as the background for a number of frames.

In the flower sequence, the motion of the objects is smooth and continuous, therefore, the objects are always segmented out well using motion information. In the tennis sequence, however, motion information fails once the player stops. In such cases, the color representation of the segments takes over, as shown in Figure 3, which results in greater temporal consistency and better segmentation than motion alone. Moreover, the segmentation is more correct in defining the object, compared to color segmentation.

Notice that since we started with a clustering of motion vectors as our initial segmentation, the first frame
consisted of a bad segmentation, but that was readily corrected within a few frames. The algorithm performed well for this sequence, for an object-based segmentation application.

6. CONCLUSION AND FUTURE WORK

This paper has presented a novel approach for fuzzy clustering of color videos. An important quality of the proposed algorithm uses spatial information of each pixel. The new method was tested on different color videos and evaluated by using various cluster validity functions. Preliminary results proved that the proposed technique well. Some methodology can be developed to evaluate the segmentation techniques on the statistical basis, so that quantifiable results can be obtained.

7. REFERENCES

[1] P. Ste and M. Domanski, “Efficient unassisted video segmentation using enhanced fast marching,” in *IEEE International Conference on Image Processing – ICIP 2003*, (Barcelona, Spain), pp. 246–253, September 14-17 2003.

[2] Sohaib Khan, Mubarak Shah, Object Based Segmentation of Video Using Color, Motion and Spatial Information, in IEEE Int'l Conference on Computer Vision and Pattern Recognition , Dec. 2001.

[3] E. Sifakis and G. Tziritas, “Moving object localization using a multi-label fast marching algorithm,” *Signal Processing: Image Communication*, vol. 16, no. 10, pp. 963–976, 2001.

[4] Ahmed MN, Yamany SM, Mohamed N, Farag AA,
Moriarty T. A modified fuzzy c-means algorithm for bias field
estimation and segmentation of MRI data. IEEE Trans Med
Imaging (2002), 21:193–199.

[5] Rosenberger, C., Chehdi, K., "Unsupervised clustering method with optimal estimation of the number of clusters : Application to image segmentation", in Proc. IEEE International Conference on Pattern Recognition (ICPR), pp. 1656-1659, vol. 1, Barcelona, September 2000,

[6] Clark MC, Hall LO, Goldg of DB, Clarke LP, Velthuizen RP, Silbiger MS. MRI segmentation using fuzzy clustering techniques. IEEE Eng Med Biol (1994) 13:730–42

[7] Trevor Darrell, Alex Pentland: Cooperative Robust Estimation Using Layers of Support. IEEE Trans. Pattern Anal Mach. Intell 17(5): 474-487 (1995)

[8] John Y. A. Wang, Edward H. Adelson: Representing moving images with layers. IEEE Transactions on Image Processing 3(5): 625-638 (1994)

[9] P.J. Burt, J.R. Bergen, R. Hingorani, R. Kolczynski, W.A. Lee, A. Leung, J. Lubin, and H. Shvaytser, “Object Tracking with a Moving Camera,” *Proc. IEEE Workshop Visual Motion*, pp. 2-12, 1989

[10] M. Irani, B. Rousso, and S. Peleg, “Detecting and tracking multiple moving objects using temporal integration,” in *European Conference on Computer Vision*, vol. 588 of *LNCS*, pp. 282–287, Springer-Verlag, 1992

[11] Thompson, W.B. and Pong, T.C. Detecting moving objects, International Journal of Computer Vision, 4(1):39-57. (January 1990).

[12] Baris David Ekinci and Muhittin Gokmen, A Ball Tracking, System for Offline Tennis Videos, Proceedings of the 1st WSEAS International Conference on VISUALIZATION, IMAGING and SIMULATION (VIS'08), Bucharest, Romania, November 7-9, 2008

[13] Cristian Molder, Mircea Boscoianu, Mihai I. Stanciu and Iulian C. Vizitiu, Improved Automatic Number Plate Recognition System, Proceedings of the 1st WSEAS International Conference on VISUALIZATION, IMAGING and SIMULATION (VIS'08), Bucharest, Romania, November 7-9, 2008

[14] Catalin Gheorghe Amza, Image Analysis based on Fuzzy Logic, Proceedings of the 1st WSEAS International Conference on VISUALIZATION, IMAGING and SIMULATION (VIS'08), Bucharest, Romania, November 7-9, 2008

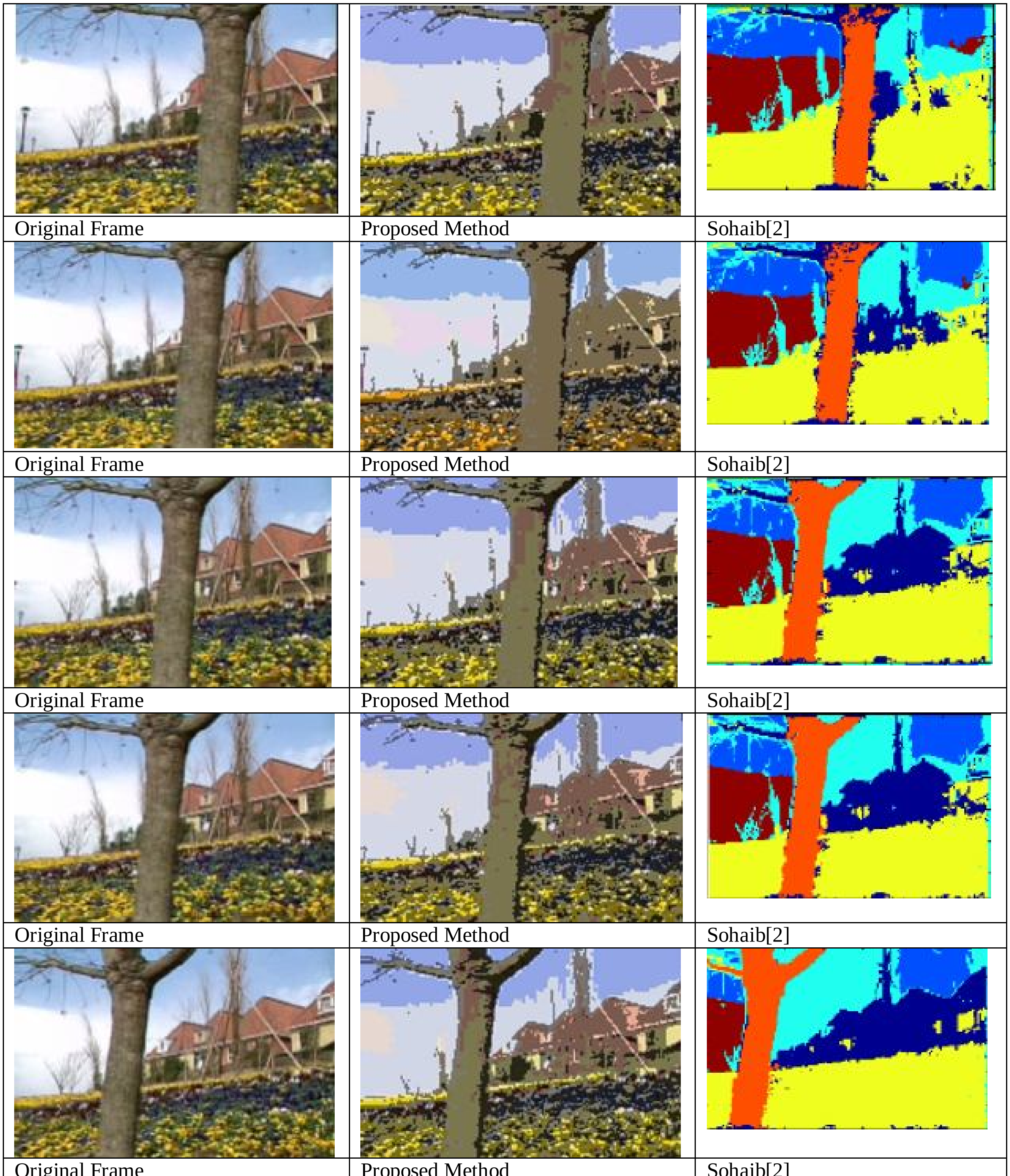


Figure 3: Results of Segmentation of Flower Sequence using both motion and color information. Each frame is segmented into different classes. Notice that proposed method has much close results to the original one. Every 4th frame is shown

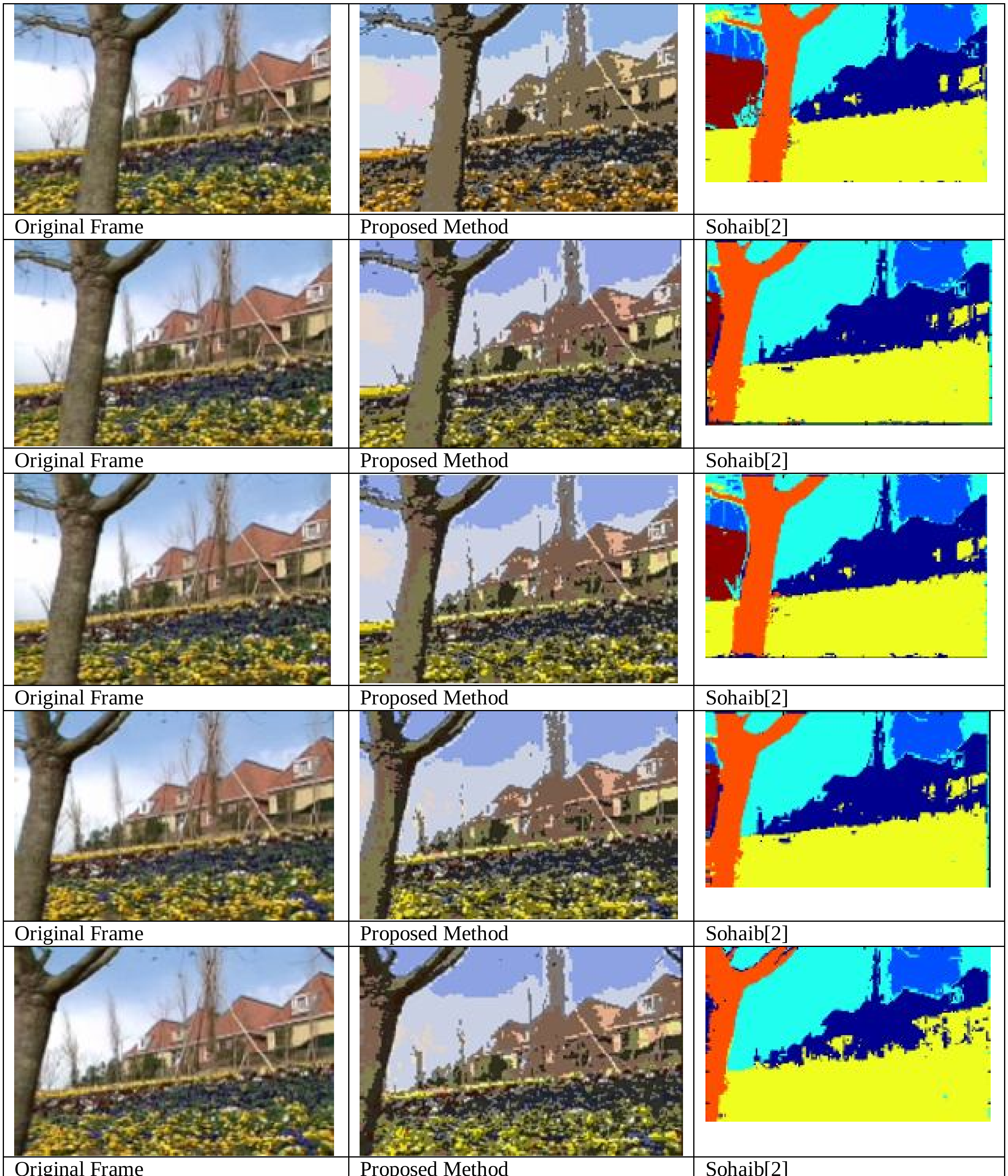


Figure 4: Results of Segmentation of Flower Sequence using both motion and color information. Each frame is segmented into different classes. Notice that proposed method has much close results to the original one. Every 4th frame is shown

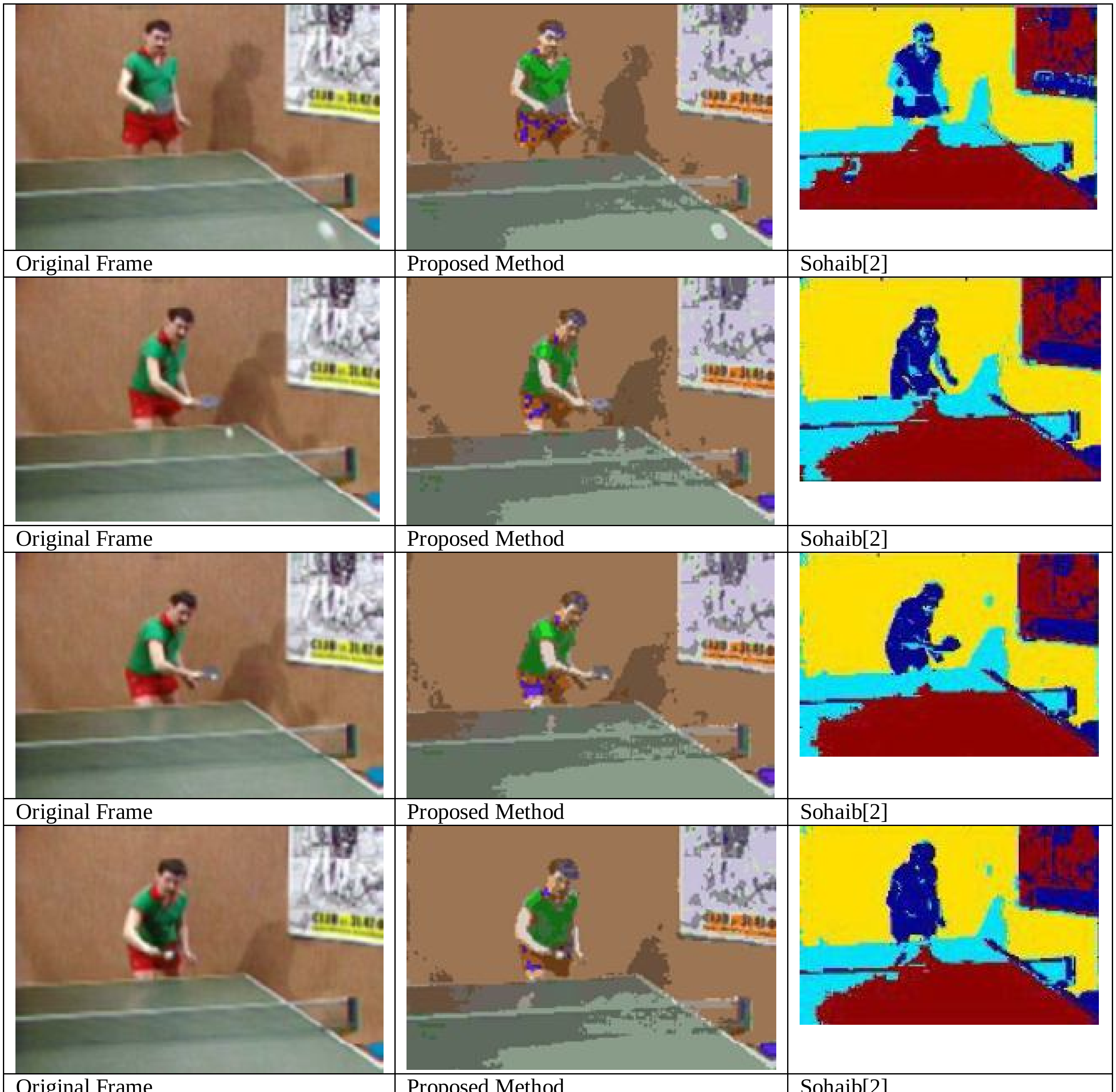


Figure 5: Results of Segmentation of Tennis Sequence using both motion and color information. Each frame is segmented into different classes. Notice that the player has inconsistent motion, often stationary, and matching with the background. However, color based segmentation takes over in this situation. Every 5th frame is shown