H-014

Automatic Object Separation of Detected Objects in Free Viewpoint Application

Houari Sabirin † Hiroshi Sankoh † Sei Naito †

1. Introduction

An application developed to improve the user experience in consuming sport videos by creating free viewpoint video has been shown in [1]. The video is made by composing the texture of player objects and blend them in a synthesized stadium or ground field. The user then can experience the sport video from any preferable angles by interactively rotates the view.

The textures of objects in the free viewpoint video are synthesized by extracting the pixels of player objects from the original videos taken by several cameras from different angles. The extraction process is performed over the original videos and it comprises of background subtraction, object detection, object separation, object tracking and finally texture pixels extraction. Currently, the aforementioned process is performed manually by human operator via the free viewpoint authoring application. It is a time consuming process where each object players in each of views of the sport video need to be extracted and annotated. The more the players appear in the videos, the more time is required to extract the objects. Nevertheless, the manual operation ensures high quality texture extraction for the final free viewpoint video.

In this paper, we aim for providing less burdensome process of synthesizing free viewpoint video by proposing an automatic method of separating objects after background subtraction. The method is performed over background-subtracted mask image in which the remaining pixels in the processed frame are the pixels of the detected player objects. The separation is performed after occlusion event has been detected and is performed upon the merged pixels of the occluded objects. The resulting separation is a clear border in the merged pixels of two (or more) occluded objects.

The detailed discussion of the proposed method is presented in Section 2 and the experimental results are shown in Section 3. Section 4 concludes this paper.

2. Proposed Method

The proposed method is applied for the object mask frames of a sport video as shown in Fig. 1. The object mask is a binary frame that contains non-zero pixels for the detected objects (throughout this paper, one group of non-zero pixels is called "object") and zero pixels for the background. For each object, we first assign its relative position $\mathbf{p}=\{p_x, p_y\}$, area $\mathbf{A}=\{a_{width}, a_{height}\}$, and its corresponding RGB data $\mathbf{I}[\mathbf{x},\mathbf{y}]$. Next the skeleton of object is generated based on thinning method [2]. This skeleton $\mathbf{S}[\mathbf{x},\mathbf{y}]$ is then used to approximate the area of the same object at the next frame, if this object is occluded with other objects. Thus the *n*-th object at frame *f* may be defined as

† KDDI R&D Laboratories, Inc.

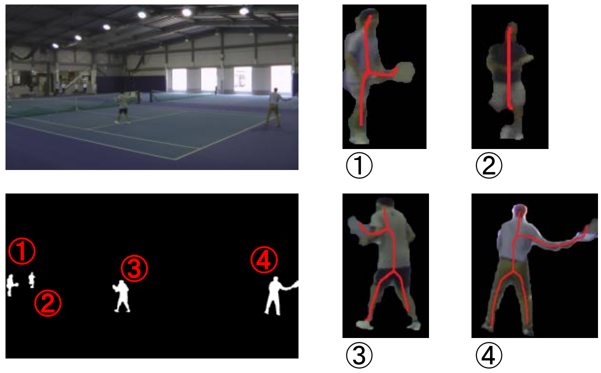

Fig. 1. Original video frame (top left), its object mask (bottom left), and the corresponding extracted objects with their skeletons (right).

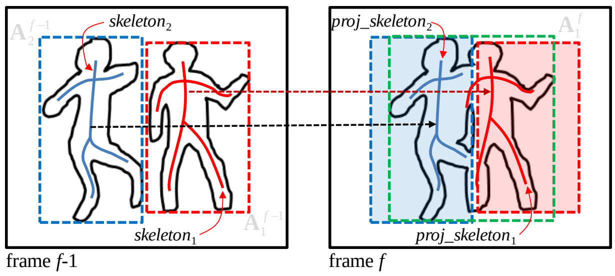

Fig. 2. Occlusion detection and skeleton markings

$$O_n^f=(\mathbf{p},\mathbf{A},\mathbf{I},\mathbf{S}). \quad (1)$$

As shown in Fig. 2, occlusion is determined by checking the overlapping area between the objects in previous frame and current frame. Let $\mathbf{A}_1^{f-1}$ and $\mathbf{A}_2^{f-1}$ be area of objects in frame *f*-1 and $\mathbf{A}_1^f$ is the area of object in frame *f*, we observe $\mathbf{\Lambda}_{11}=\mathbf{A}_1^f\cap\mathbf{A}_1^{f-1}$ and $\mathbf{\Lambda}_{12}=\mathbf{A}_1^f\cap\mathbf{A}_2^{f-1}$. An occlusion between O_1^{f-1} and O_2^{f-1} is then determined when

$$\mathbf{\Lambda}_{11}=\mathbf{\Lambda}_{12}\geq 75\%. \quad (2)$$

If this condition is satisfied (i.e. an occlusion is occurred), the separation is performed to find the boundary between the occluded objects.

Separation is performed by utilizing watershed algorithm [3]. Since the watershed algorithm requires segmentation markers to determine which parts of an area of image to be separated, our separation method focuses on automatically generating such markers. Watershed algorithm also relies on the saliency of colors to be segmented. Given the segmentation markers, we can generally separate objects according to their location in previous frame. Let $\mathbf{S}_1^{f-1}$ and $\mathbf{S}_2^{f-1}$ be the skeletons of two already separated objects in frame *f*-1, the skeleton for the overlapping object in frame *f* that satisfy equation (2) is determined as $\tilde{\mathbf{S}}_1^f=\{\mathbf{S}_1^{f-1},\mathbf{S}_2^{f-1}\}$. Based on this set of projected skeletons,

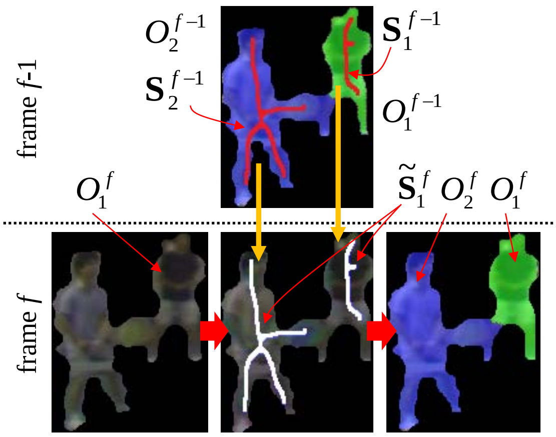

Fig. 3. Projecting the skeletons from the previous frame with separated objects as references for separation in current frame.

the watershed algorithm is then performed over RGB data $\mathbf{I}$ to obtain the new objects O_1^f and O_2^f.

Nevertheless, in many cases, the overlapping area of occluded objects tends to grow larger and fully covering the objects behind it. In this case, the approximation of segmentation markers may not be sufficient to separate the pixels of the objects. To enhance the separation for watershed algorithm, color enhancement based on histogram equalization is performed.

From the experiments, it can be determined that XoYoR color space can enhance the saliency of colors. The XoYoR color space is a mix of XY channels and R channel from XYZ and RGB color spaces, respectively. Additionally, to enhance the area between two different colors (such as the borders between two objects), histogram equalization for one selected channel in a color space is needed. However, currently there is no specific color space that provides general optimal solution for different video sequences. Different color spaces for histogram equalization are utilized to find one that most fit with the sequence being tested. Combination of several histogram equalizations from different color spaces is also utilized to enhance the saliency. Figure 3 illustrates the projection of the skeletons of the separated objects in previous frame to be used as segmentation markers for object separation in current frame in *Tennis* sequence. Here, the histogram of hue channel of HLS color space is enhanced for the watershed algorithm.

3. Experimental Results

We tested our method in several sport sequences, namely *Baseball*, *Badminton*, *Tennis* and *Futsal* sequences. Each sequence has more than two views taken from different angles. The proposed method is applied to all views in each sequence by adjusting the histogram equalization for luminance channel of a color space that produces the best qualitative separation results. Table 1 shows the composition of color enhancement applied to the different sequence. The composition is determined by observing the characteristics of objects and colors in each sequence.

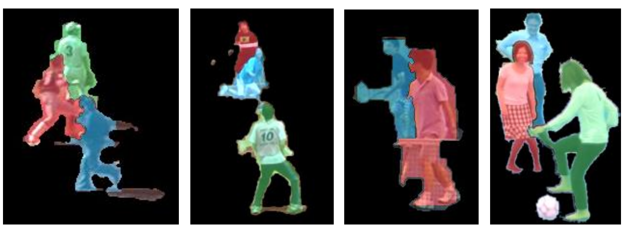

Fig. 4. Separation of occluded objects in (from left to right) *Baseball*, *Badminton*, *Tennis*, and *Futsal* sequences

Table 1 Composition of color channels for histogram equalization

Sequence	**Color composition [selected channel]**
Baseball	YCrCb[Y]
Badminton	XYZ[Z] + CIELab[L]
Tennis	HLS[H]
Futsal	XYZ[Z] + CIELab[L]

Figure 4 shows the results of separating occluded objects in *Baseball*, *Badminton*, *Tennis*, and *Futsal*, respectively. *Baseball* and *Badminton* sequences have small sized objects relative to the frame size, while *Tennis* sequence has medium sized objects and *Futsal* sequence has large sized objects. The objects in *Badminton* and *Baseball* sequences are relatively easier to be separated due to the color of shirts worn by the players. In *Futsal* sequence, the players are wearing almost the same colors of attire (light-colored shirts and dark-colored pants/skirt). In *Tennis* sequence, the colors of the players' shirt are unclear for the objects that are far from the camera.

Some cases of incorrect separation that commonly occur include incorrect approximation of its skeleton from previous frame, incorrect separation due to color similarity and the insufficiency of saliency between two or more unclear colors.

4. Conclusions

We have shown in this paper a novel method of automatic generation of segmentation markers for watershed algorithm. The method is utilized to separate the occluded objects in sport sequences to generate free viewpoint video. Although the problems with insufficient color saliency and color uniformity may occur, the overall accuracy is relatively high as shown in the experimental results. Further observation in enhancement of color from video data will be conducted to improve the separation results.

References

[1] H. Sankoh and S. Naito, "Free-viewpoint Video Rendering in Large Outdoor Space such as Soccer Stadium based on Object Extraction and Tracking Technology," The Journal of The Institute of Image Information and Television Engineers (ITE), Vol. 68, No. 3, pp. J125-J134, 2014.

[2] T. Y. Zhang and C. Y. Shuen, "A fast parallel algorithm for thinning digital patterns," Communications of the ACM Magazine, Vol. 27, No. 3, pp. 236-239, March 1984.

[3] F. Meyer, "Color Image Segmentation," in Proc. Int. Conf. on Image Processing, pp. 303-306, 1992.