Video Retrieval based on User-Specified Deformation

Yuta Kawate[†] Makoto Okabe[†, *] Rikio Onai[†] Hiromi Hirano[‡] Masahiro Sanjo[‡]
[†] The University of Electro-Communications [*] JST PRESTO [‡] Rakuten, Inc. Rakuten Institute of Technology

1. Introduction

Let's assume that we are now watching the scene of Fig. 1-a, where a red racing car is running from left to right. At the same time, we feel like watching another scene like Fig. 1-b, where a similar car is running toward us. To find such a desired scene, we usually push the forward and rewind buttons or move the play bar. If we cannot find the scene in the currently watching video, we have to go to a video sharing service like YouTube to further search for it. However, these are tedious tasks. Goldman et al. propose a method to navigate a single video [2008], but we want to find a scene not only in a single video but also a different video from the currently watching one.

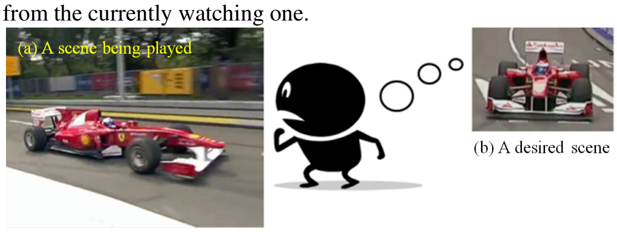

Figure 1. Our motivation.

We propose a novel user interaction for quickly and easily playing the desired scene. Our idea is based on the observation that the car in the desired scene (Fig. 1-b) is the rotated version of the car in the currently watching scene (Fig. 1-a). We then allow the user to specify this deformation. In Fig. 2-a, the user draws the two green arrows intending to move the front and back spoilers to tips of the arrows. Then, our system immediately retrieves the desired scene, where the location of each spoiler corresponds to the tip of the arrow (Fig. 2-b).

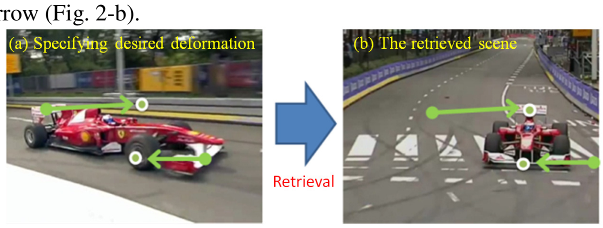

Figure 2. The proposed user interaction.

2. Our Approach

It is difficult to directly find the desired scene (Fig. 2-b), using the currently watching scene (Fig. 2-a) as a query of a standard image or video retrieval technique. For example, we extract SIFT features in each frame of the video and compute their matching between the currently watching scene and the other frames. As shown in Fig. 3, we successfully find many consistent matches with a similar looking scene: it is left side right but the posture of the car is similar. However, there is no consistent match between the cars in the right case.

However, it is interesting to note that the bottom two frames in Fig. 3 exist in the same video sequence: the bottom-left frame comes several seconds after the bottom-right frame, i.e., the car is coming toward us and then turning to the left. We can find the desired scene by rewinding the video from the frame of Fig. 3-left. During the rewinding process, we track the front and back spoilers, since they are specified by the user as tails of the arrows. We stop to rewind the video, when each tracker comes near to the tip of the corresponding arrow. Then, the desired scene has come.

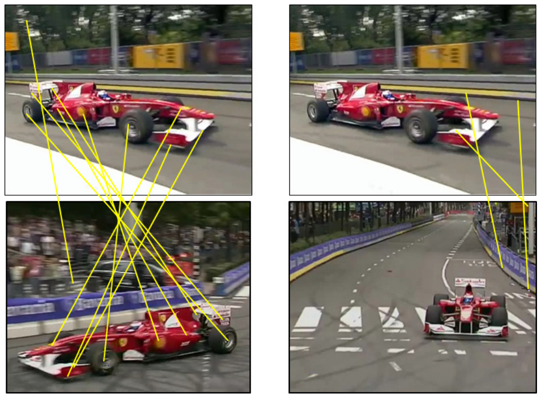

Figure 3. Video retrieval using SIFT matching. Left: We find many matches. Right: We cannot find any consistent match.

In conclusion, our algorithm consists of two processes: first, we find a frame that looks similar to the query frame using SIFT matching, and second, we automatically forward and rewind the video from the similar looking frame to find the user-desired scene. SIFT matching is efficiently performed using kd-tree algorithm. To efficiently find the desired scene by forwarding and rewinding the video, we use Particle Video algorithm [P. Sand et al. 2008] for motion tracking.

3. Results

We experiment our method on two videos downloaded from YouTube: one is a video of racing cars, and the other is of running horses. The duration of them are 3,718 and 4,450 frames, i.e., several minutes long. We spend about four seconds in average for retrieval process using an Intel i7 2.8GHz processor. The input and retrieved scenes are shown in Fig. 4.

Figure 4. Retrieval results.

References

D. B. GOLDMAN, C. GONTERMAN, B. CURLESS, D. SALESIN, AND S. M. SEITZ. 2008. Video object annotation, navigation, and composition. In *Proc. of UIST*, 3-12.

P. SAND AND S. TELLER. 2008. Particle Video: Long-range motion estimation using point trajectories. *IJCV*, 80, 1, pp 72-91.

SIGGRAPH 2012, Los Angeles, California, August 5 – 9, 2012.
ISBN 978-1-4503-1435-0/12/0008