Automatic Composition of an Informative Wide-View Image from Video

Hitoshi Habe Shota Makiyama Masatsugu Kidode
Graduate School of Information Science
Nara Institute of Science and Technology
Ikoma, Nara, Japan
E-mail: habe@computer.org

Abstract—We describe a method for generating an informative wide-view image using images captured by a moving camera. The generated image allows for events in the scene observed by the camera to be understood easily. Our method does not use 3D shape information explicitly. Instead, it employs the trajectory of feature points across multiple images and generates a composite image by taking into account the distribution of the trajectories of the feature points.

Keywords—video volume; wide-view image; feature points; informative image

I. INTRODUCTION

When we capture images of a complex scene by using a moving camera to scan the scene, we need to observe a long sequence of images in order to understand the details of the scene. If a single image composed from an image sequence could be used to provide important information about the scene, it would save time and effort and could be applied to various scenarios including surveillance, indexing, and summarization.

The simplest method to composite a single image from an image sequence is image mosaicing [1], [2]. In this method, each image in the sequence is first aligned, and then, a single image is composited from the sequence. Although this method works well for a scene containing simple objects such as planar objects, it fails when a scene contains complex objects such as convex and/or concave objects; the alignment would fail because of appearance variation and some occluded regions that would not appear on the composite image.

All the information included in the image sequence can be represented by the video volume [3], [4], as shown in Fig. 1. The video volume is a three-dimensional space having three axes, x, y, and t. Each image in the sequence corresponds to a slice on an x − y plane. Using this representation, we can synopsize the scene in an image or video.

For example, we can generate a synopsized video by removing less informative portions from the volume and by then reconstructing a video from the volume [5], [6]. In addition, slicing the volume using a dynamically deforming surface produces attractive videos with effects such as slowing down and speeding up [7].

The video volume can also be used for compositing a single image by slicing the volume using a surface having an arbitrary shape. Other uses include interactive editing [3], optimization of a slicing surface with regard to the consistency in resultant images [4], and finding out the best slice using the scene depth [8].

In this paper, we describe a method that automatically composites an informative image from a video sequence. Here, an informative image is one that includes the surface of objects in the scene to the greatest extent possible even when the surface is not visible in some images in the sequence due to occlusion. In contrast with the aforementioned work, we define an objective function based on feature point tracking using a SIFT descriptor [9] to detect and show informative regions.

Among the aforementioned works, [3] and [4] do not take into account the content of the video. They require the user’s assistance for identifying important regions. Using the scene depth, as in [8], would enable us to achieve our abovementioned goal. However, it is difficult to acquire the depth information in a real scene. Therefore, we introduced a method that uses only the 2D information in images and composites a single image based on the feature point tracking of 2D images.

II. INFORMATIVE IMAGE GENERATION USING VIDEO VOLUME

The video volume has been used in some studies to generate a new image or video from a video [3], [4], [8]. This section first describes the concept of image composition using the video volume. Then, we discuss some technical issues involved in compositing an informative image using
the video volume, and we propose our idea for solving these issues.

A. Image Composition using Video Volume

Slicing the volume using a surface provides a single image containing information about the input sequence. As shown in Fig. 2, the choice of the slice determines how the resultant images look. Our problem here is to choose a suitable slice from the volume that provides an informative image.

B. Requirements on Slice

When one chooses a slice for image composition, the slice has to satisfy two requirements. First, it has to be chosen while maintaining the pixel order of the original images in a composite image. Second, it should provide an informative image, as mentioned earlier. Hereafter, we call these requirements the pixel order requirement and informative region requirement, respectively.

Fig. 3 illustrates the necessity for introducing the pixel order requirement. In (a), the points a and a’ correspond to the same points in the scene, as do b-b’ and c-c’. In the video volume, these points are shown as illustrated in (b) and (c); the flow of each point is determined by both the camera motion and the depth. Here, the slice has to be drawn such that it crosses all the flows from the same direction, for example, from right to left when one sees the origin of the flow. If these requirements are not satisfied, as in the area indicated by the dotted circle in Fig. 3(c), the order of the pixels in the generated image becomes inconsistent, as shown in Fig. 4, where the dashed circles indicate the area having a pixel order that differs from those of other areas.

Fig. 5 illustrates another video volume. The right- (a) and left-hand sides (b) of the box do not appear in all the images in the sequence. If the slice does not pass the area in the video volume corresponding to the partially occluded areas (a’ and b’ in the figure), these areas are not shown in the resultant image. Our goal in this paper was to choose a slice so that partially occluded areas appear in the resultant image to the greatest extent possible. This is the informative region requirement. This enables us to provide an informative image that enables observers to see what is in the scene.

Here, it should be noted that the pixel order requirement is almost equivalent to the constraint used by Wexler et al. [4]. In contrast, our principal contribution is to consider the image contents by defining the informative region requirement.

III. Our Method

This section describes our proposed method. It should be noted that in this study, we restricted ourselves to slices that form a ruled surface in a video volume. As discussed in [4], this is a common scenario corresponding to a camera moving parallel to the ground plane. Hence, slice optimization is equivalent to finding an optimal path on the $x-t$ plane in the video volume. We call the path the slice path.
A. Algorithm

The algorithm for finding an appropriate slice path is as follows:

1) Generate a video volume by accumulating each image in the input sequence.
2) Extract feature points and track the points across the frames using a SIFT descriptor.
3) Find an appropriate slice path by carrying out optimization using an objective function.

The main step in this algorithm is Step 3, especially the design of the objective function that satisfies the requirements mentioned in Sec. II-B. The details of the same are described in the next section.

B. Slice Optimization for Informative Image

As discussed in Sec. II-B, we assume that an informative image implies that we can see as many object surfaces in the image as possible. Via feature point tracking, as described in Step 2, we obtain several trajectories; each trajectory corresponds to a feature point on the surface of an object. If a slice path passes a trajectory, the corresponding feature point will appear in the resultant image. Therefore, to satisfy the informative region requirement, we can define an objective function as

\[ E(p) = \sum_{p_i} T_i, \]

where \( p = \{p_1, p_2, \ldots, p_N\} \) denotes a slice path on an \( x-t \) plane; \( p_i \), a small portion on the slice path; and \( T_i \), the number of trajectories that \( p_i \) passes. Based on the assumption, if \( E(p) \) takes a higher value, the slice path \( p \) provides a more informative image.

With regard to the pixel order requirement, once a slice path does not satisfy this requirement, the resultant composite image becomes unnatural, as shown in Fig. 4. Therefore, instead of an objective function, we impose a constraint in the optimization process so that a resultant slice path satisfies the pixel order requirement.

In summary, we formalize the optimization problem as follows:

\[ p = \arg \min_p E(p), \quad \text{s.t.} \quad p \in \mathcal{P}, \]

where \( \mathcal{P} \) is a set including paths that satisfy the pixel order requirement.

It is difficult to determine a value of \( p \) that satisfies Equation (2) from all possible solutions, i.e. slice paths, because of the large search space; for example, the slice path can be drawn on a \( 640 \times 377 \) array for Experiment 1, as shown in Figure 7. To make the optimization simple and feasible, we divided the \( x-t \) plane into blocks and performed an optimization on the resolution of the blocks by applying the Bellman-Ford algorithm [10]; this algorithm searches a path that has the smallest sum of weights along it.

IV. Experimental Results

We conducted an experiment to demonstrate the effectiveness of our method. Four videos were captured using a camera moving parallel to the ground plane. The videos and results are shown in Figures 7, 8, 9, and 10. Each image in the videos has a resolution of \( 640 \times 480 \), and the videos contain 377, 391, 306, and 446 images, respectively.

As mentioned in Section III, we first detected the feature points in images and found their correspondence along the time dimension \( t \). The short lines in Figure 6 indicate the results of feature point tracking for each video. Then, we found a suitable slice path based on the formulation of the optimization problem introduced in Sec. III-B. The blue lines shown in Figure 6 indicate the respective slice paths.

By applying the slice path to the video volume, we obtained the composite image shown at the bottom of Figures 7, 8, 9, and 10. Although the areas indicated by dashed circles in the resultant images did not appear in all the input images, they appeared properly in the composite image. This implies that our method did not leave out this important information.

However, the most noticeable drawback in the composite images is shape distortions. All objects in the composite images are grossly distorted. It is quite natural for more shape distortions to appear with a greater number of surfaces of scene objects in the composite image. This problem is a type of trade-off involved in generating an informative image.

Additionally, the computational efficiency remains a problem. As discussed in Section III, we restricted the feature space in order to obtain results efficiently; actually, we obtained the results in several hours. However, this led to the presence of some artifacts in the results; because the slice path was not smooth, the composite images had unnatural boundaries. In future work, attempts must be made to develop a more efficient and effective optimization method.

Figure 6. Feature tracking and slice path detection results
V. SUMMARY

We described a method for generating an informative wide-view image from a video captured using a moving camera. The method does not recover and utilize 3D information in the scene explicitly. Instead, it uses a video volume representation for image composition. A slice path is chosen using the trajectory of feature points so that the path passes through as many trajectories as possible.

To the best of our knowledge, this is the first work that reports a method for automatically generating a composite image from a video; existing methods require the user’s assistance. Although the results presented in this paper are preliminary, they demonstrate that the method is promising.

We believe that further studies will improve the functionality and performance of this method. Specifically, as discussed in Section IV, our future work will attempt to develop an efficient and effective optimization method for Step 3 and develop an extension to a slice having an arbitrary shape.

ACKNOWLEDGMENT
The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

REFERENCES