Automated Feature Points Management for Video Mosaic Construction

Jing Li[1] , Quan Pan[1], Tao Yang[1], Stan.Z. Li[2]
[1]*College of Automatic Control, Northwestern Polytechnical University, Xi'an, P.R . China , 710072*
[2]*National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, P.R.China, 100080*
jinglinwpu@163.com, quanpan@nwpu.edu.cn，yangtaonwpu@163.com，szli@nlpr.ia.ac.cn

Abstract

This paper presents a two level automatic feature points management method for constructing a seamless entire panorama from video sequence. In the first level, through fusing the number of tracked feature points and the estimated ratio of lost information of the mosaicing image, a feature point quantity management module is developed to select the key frames. In the second level, a feature points quality management technique is used to choose the key points for mosaicing. This module includes a coarse-to-fine method with two steps: (1) Feature points quality based key point subset creation, and (2) Multi-resolution based key point selection. The main contribution of the algorithm is that it is able to achieve robust and fast mosaicing result while maintain the most valuable information of the scene. Experiments are performed using video sequences under different conditions. The results show that the proposed algorithm could achieve robust and efficient video mosaic image.

1. Introduction[1]

Mosaic Construction is an active area of research in computer vision and it has various applications such as satellite photographs, video surveillance, stabilization, compression, virtual environments, virtual travels and 3D world scene medical imaging[1,2].

Numerous techniques have been approved for image mosaicing which can be classified broadly into: direct methods [3,4] and feature-based methods [5,6]. Direct methods use information of all pixels and discover parameter set through an iterative process to minimize the sum of squared difference (SSD). However, these methods require good initial values for the parameters of the transform. If they are not corresponding to physical movements of the camera, it is very difficult to evaluate these parameters value [7]. On the other hand, feature-based methods have a common difficulty in acquiring and tracking of image features. Therefore, many researchers have paid much attention to how get good features and track more accurately.

In this paper, we care for the feature-based methods for video mosaic construction. Here we do not specify how to select and track feature points. The problem we address in this paper comes from two aspects:

(1) Since a video contains significant redundancy, so that not all frams are required to create a mosaic. In order to meet the real time demand in many systems, only some key images are selected to create a mosaic. So how to decide these key images automatically is an important problem.

(2) Feature-based methods always have the assumption that the corresponding points are tracked correctly enough, however, if error corresponding points appearance, or the tracking result is not very precise, it will deeply influence the performance of mosaic. Furthermore, given that all correspondences are correct enough, choose different pairs of feature points may result in different accurate of mosaic. So how to choose the most suitable pairs of feature points is also a problem.

To solve the problems above, a two level automatic feature points management method for video mosaic is presented to select key frames and points for final mosaic. A unique character of this method is that it achieves highly accurate results automatically.

The remainder is organized as follows. Section 2 presents an outline of the algorithm. Section 3 describes the automated feature points management in details, including quantity management and quality management. Various experimental results are presented in Section 4 to illustrate the performance of the feature points management method.

[1] The work presented in this paper was sponsored by the Foundation of National Laboratory of Pattern Recognition and National Natural Science Foundation of China (60172037)

2. Outline of the proposed algorithm

The video mosaicing algorithm consists of five parts (shown in Figure 1): (1) initialization (2) feature points tracking (3) feature points management (4) image mosaicing, and (5) feature points reselection. Two modules are included in part three: feature points quantity management and feature points quality management. In the first module, by estimating the number of tracked feature points and the ratio of lost information of the mosaic image, the key frames are selected. Then, in the second module, quantity based rules will be designed to choose the key points for mosaicing. In part four many mosaicing models can be chosen according to various conditions. At last, in part five, new feature points are reselected from the current key frame for further feature points tracking.

3. Automated Feature Points Management

3.1. Initialization and feature points tracking

First, we select N feature points on the initial frame $F(k)$ of the video to get the feature points vector $f(k)=(f_1(k), f_2(k)\ldots f_N(k))$, $f_i(k)$ contains the position of the *ith* feature point which defined as $f_i(k)=\{x,y\}$ at time k . Meanwhile, $F(k)$ is taken as the current reference frame $F_{\mathrm{Re}f}$. Second, in part two, a points tracking method mentioned in [3] are used here. After λ frames, the state vector of the feature points at time $k+\lambda$ is denoted as $T(k+\lambda)=(T_1(k+\lambda),\ldots,T_N(k+\lambda))$,

$$T_i(k+\lambda)=\begin{cases}1 & tracked \\ 0 & otherwise\end{cases} \quad i=1\ldots N \tag{1}$$

The state vector $T(k+\lambda)$ is send to the third part and be managed to search the key frames for mosaicing.

3.2. Quantity Management

This section will describe the feature points quantity management in detail. The main purpose of this module is to search key frames from the original input video for video frames are typically 30 fps and contains significant redundancy.

In order to select good key frames, several factors have to be considered. Firstly, enough pairs of feature points must be preserved to make sure of the precision and accuracy of mosaic result; Secondly, the overlapping area between relative key frames should be large enough to avoid the lost of scene information.

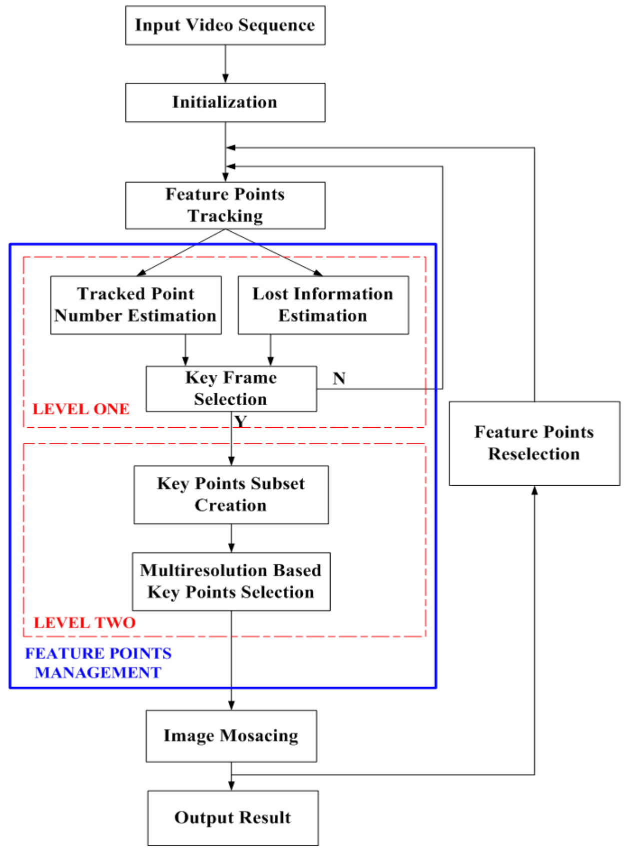

Fig 1. The block diagram of the algorithm

So in this paper, we design two parameters to define good key frames: tracked points number and ratio of lost information.

As described in section 3.1, we take the $F(k)$ as the current reference image $F_{\mathrm{Re}f}$, and keep on tracking the feature points vector $f(k)=(f_1(k),\ldots,f_N(k))$ frame-to-frame, according to the state vector $T(k+\lambda)$, tracked feature points number at time $k+\lambda$ is denoted as $M(k+\lambda)$

$$M(k+\lambda)=\sum_{i=1}^{N} T_i(k+\lambda) \tag{2}$$

M is started from N, and with the frame interval enhance, it will decrease to an threshold which can be designed beforehand. In this paper, we consider this threshold $TH1$ as fifty percent of the initial number N. That means, when M decreased to a relatively low level, we think that these two frames $F(\mathrm{k})$ and $F(\mathrm{k}+\lambda)$ include enough different information and would be the key frames for mosaicing.

Furthermore, considering that if some strong feature points exist and they are almost matched very well during many frames. The situation above will be very difficult to reach, thus the interval between $F(k)$ and $F(k+\lambda)$ are very large and lost too much information between these two images. Therefore, in order to ensure key frames remain enough information, we give an upper limitation $TH2$ for the frame interval. If the frame interval between the two key frames is larger than $TH2$, we should select these two frames as the key frames even if the M is higher than $TH1$.

In this paper, we also proposed a general idea to identify the upper limitation $TH2$ in different situations. That is the overlapping area between two key frames should remain enough information for this scene; meanwhile the lost information must be smaller than a limit. Given there are some successive frames from a video: $F(m)\ldots F(m+L)$. The Overlap_Area and Lost_Area for $F(m)$ and $F(m+L)$ are defined as follows:

$$Overlap\_Area = \frac{Area(F(m) \cap F(m+L))}{Area(F(m))} \tag{3}$$

$$Lost\_Area = \frac{Area(F(m) \cup F(m+1) \cup \ldots \cup F(m+L)) - Area(F(m) \cup F(m+L))}{Area(F(m))} \tag{4}$$

Then formulation (5) is designed to realize this idea.

$$\begin{cases} Overlap\_Area > P_{overlap} \\ Lost\_Area < P_{lost} \end{cases} \tag{5}$$

Where $P_{overlap}$ is used to make sure the Overlap_Area is enough large is often set to 30%. At the same time, the Lost_Area should not be too large to influence lot the final mosaic, so another parameter P_{lost} is often been set to 10%. If the formulation (5) is met, tracking will be continue on next frame . Once one of the two inequations has been broken, the frame interval between $F(m)\ldots F(m+L)$ is the upper limitation $TH2$. This parameter $TH2$ can be calculated offline.

Therefore, the quality management module can be described as following: tracking initially N points selected on $F(k)$ frame-to-frame until tracked number M reaches the $TH2$ or the frame interval λ reaches the upper limitation $TH2$, thus the current frame $F(k+\lambda)$ for mosaicing has been selected. The feature points vector $f=(f_1,f_2\ldots f_N)$ has change to be the $f'=(f'_1,f'_2,\ldots f'_M)$ ($M \le N$) which is the tracked feature points on $F(k+\lambda)$. And we obtain the M pairs corresponding points in all $f_P=\{(f_j,f'_j)\,|_{j=1\ldots M}\}$.

3.3. Quality Management

As the 3.2 section chooses the key frames and corresponding tracked feature pairs above, father, to ensure the accuracy of the final mosaic result, some evaluation methods are used here to find the key points instead of randomly selection.

In this subsection, we present a coarse-to-fine method which includes two steps to select the key points: (1) feature points quality based key point subset creation, and (2) multi-resolution based key points selection.

(1) First we create a subset points based on the feature points quality Q and the distance D between two key points. The subset feature points f_C which include H pairs meet the conditions that both Q and D are large enough. According to these two limitations we obtain the key point subset which includes those high quality and large distance feature points.

(2) Second we search the most suitable key points in f_C under the multi-resolution frame. As suppose above that all the frames are captured on the same plane, so we adopt the affine transform model for mosaicing, and two pairs of key points are needed.

1. Down sampling the two images F_{Ref} and $F(k+\lambda)$;
2. Randomly sample two pairs among the candidate pairs f_C and iterate the $s_f=\infty$;
3. Compute the transform matrix T determined by these two pairs ;
4. Finding the overlap area $Over\_F_{\mathrm{Ref}}$, $Over\_F(k+\lambda)$ in mosaicing image by using this matrix T, Where $Over\_F_{\mathrm{Ref}}$, $Over\_F(k+\lambda)$ are from the two original images $F_{\mathrm{Ref}}, F(k+\lambda)$ separately;
5. Calculate the difference between $Over\_F_{\mathrm{Ref}}$ and $Over\_F(k+\lambda)$. Here , we use the simple function as following:

$$difference(X,Y)=\frac{1}{W}\sum|X-Y|^2 \tag{6}$$

where W is the sum number of pixels. Therefore,

$$s = difference\,(Over\_F_{\mathrm{Ref}}, Over\_F(k+\lambda)) \tag{7}$$

6. If $s<s_f$, update s_f : $s_f \leftarrow s$

Repeat the above computations until s_f reaches its minimum. Then regarding the two pairs as the most suitable key points vector f_S for mosaicing.

3.4 Mosaicing and Feature Point Reselection

Using $F_{\mathrm{Re}f}$, $F(k+\lambda)$ and $f_S = \{(f_a, f_a'), (f_b f_b')\}$ to do mosaic construction and obtain the most accurate mosaic image including the context of Frame $F_{\mathrm{Re}f}$ to Frame $F(k+\lambda)$. Updating the current key frame F_K to the new reference image $F_{\mathrm{Re}f}$ and reselect N feature points on it again.

4. Experiment Results

The technique for solving the automated video mosaicing problem has been tested with various sets of images. The video images are captured by Sony EVI-D100 at 25fps for the image size of 320x240 (24 bits per pixel). The algorithm is tested with complex outdoor video sequences and achieves satisfied results. The follows represent the results.

Figure2 shows four decrease curves of feature points number, the threshold $TH1$ is shown in Figure2 with a broken line. Once the tracked point number is less than $TH1$, the reselection of feature points included in quantity management module is active. Through Figure2 we can see that with the frame interval increase, the number of tracked points decrease rapidly. Figure3 displays the different s with various combinations of points in f_C. Note that the #9 has the lowest difference and its mosaic result with best performance. Figure4 shows a mosaic result of an outdoor sequence. This sequence contains 635 frames, and the method automatic select eleven key frames from it to build the final image. Note that the experiment result is robust and accuracy with little lost information about the scene.

5. Conclusion

An automatic feature points management for video mosaic construction algorithm has been proposed. By analyzing the quantity and quality of feature points, the key frames and the most efficient and effective feature points can be selected automatically to get the large view mosaic image for many applications such as surveillance etc.

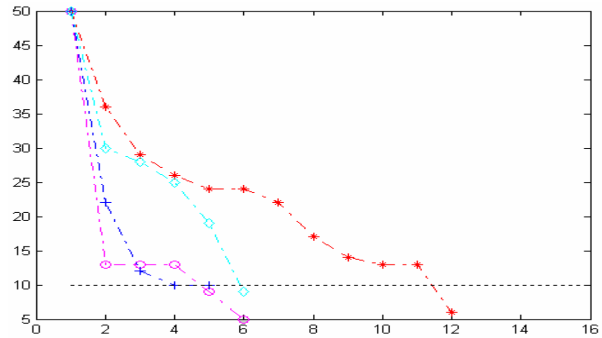

Figure 2. Decrease curve of feature points number

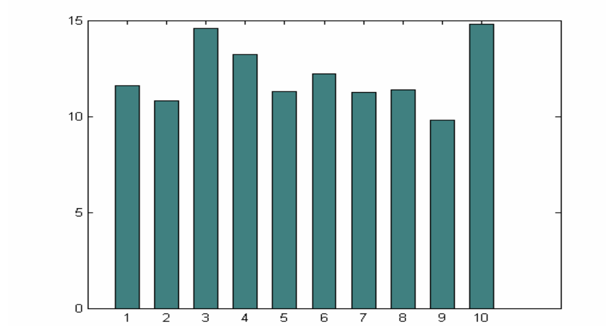

Figure 3. Difference map under various feature points selection.

Figure 4. Mosaicing result from an input video sequence. 11 key frames are selected from 635 frames of the video sequence.

6. Reference

[1] R. Kumar, H. Sawhney, S. Samarasekera, "Aerial Video Surveillance and Exploitation", *IEEE proceedings*, vol. 89, no. 10, October 2001, pp. 1518-1539.

[2] R. Szeliski, H-Y. Shum ,"Creating full view panoramic image mosaics and environment maps", *Proceedings of the SIGGRAPH'97*, CA, August,1997, pp:251–258.

[3] I. Zoghlami, O. Faugeras, R. Deriche, "Using geometric corners to build a 2D mosaic from a set of images", *Proceedings of the IEEE Conference on CVPR,* Puerto Rico June, 1997, pp: 420–425.

[4] H.Shum,, R.Szeliski,, "Construction and refinement of panoramic mosaics with global and local alignment", *IEEE Int'l Conf. Computer Vision*, 1998, pp. 953-958.

[5] R Hartley, A.Zisserman, "Multiple View Geometry", *Cambridge University Press*, UK, 2000.

[6] D.Capel, A.Zisserman, "Automated mosaicing with super-resolution zoom", *Computer Vision and Pattern Recognition*, 23-25 Jun ,1998, pp: 885-891.

[7] C.Tat-Jen, R.Clipolla, "A Statistical Framework for Long-Range Feature Matching in Uncalibrated Image Mosaicing", *Proceedings of the IEEE. Conference on CVPR,* CA, June. 23-25, 1998, pp.442-447.