An Improved Algorithm on Image Stitching based on SIFT features

Jing Xing, Zhenjiang Miao
Institute of Information Science, Beijing JiaoTong University, Beijing 100044, P.R. China
xingjing2006@gamil.com

Abstract

IBR (Image-based Rendering) is an important technology in VR (Virtual Reality). Panoramic image stitching used to create virtual environment for many applications is a key technology for IBR, and lots of stitching algorithms are developed in recent years. In this paper, we introduce an algorithm based on SIFT features to stitch panoramic images. At last, some improvements are made that we stitch the two images by using the coordinates of the image after transformation instead of interpolation; in this way complex algorithm and unnecessary calculation are avoided. Through lots of analyses and experiments we find that it is very effective. Because this method is not sensitive to ordering, orientation, scaling and illumination, its stitching precision is much higher than many other methods.

1. Introduction

With the widely application of Virtual Reality technology, the reality of virtual environment is required much higher. Because of the limitation of geometry-based Rendering, Image-based Rendering (IBR) is becoming more and more important. Panoramic image stitching is a technique to merge a sequence of images with limited overlapping area into one blended picture. The automatic construction of image stitching is an active research area in Virtual Reality (VR) which is an important developing direction of simulation technology. So satisfying image stitching results are crucial in the construction of virtual environment, which means a natural transition from one image to another, both in structure and intensity within and possibly beyond the overlapping region.

In the research literature, methods for automatic matching fall broadly into two categories: direct and feature-based. Direct methods attempt to iteratively estimate the camera parameters by minimizing an error function based on the intensity difference in the area of overlap. Direct methods have the advantages that they use all of the available data and hence can provide very accurate registration, but they depend on the illumination.

Feature-based methods begin by establishing correspondences between points, lines or other geometrical entities. Recently there has been great progress in the use of invariant features for object recognition and matching. These features can be found more repeatedly and matched more reliably than traditional methods. Because feature-based methods have higher accuracy, some image stitching methods are introduced and implemented in this paper. Based on our discussion, some satisfying stitching results are obtained using the SIFT features-based method.

2. Algorithms details and experiment results

It's a key to match features in the images stitching. So the result of image stitching will be good if the interest points are found correctly. Among the local descriptors compared, SIFT features generally perform the best [1] [4] [7]. Because of the unreliability of the many algorithm such as the algorithm based on area and based on feature pattern [2], the algorithm based on SIFT features is introduced. The algorithm of image stitching based on

0-7695-2882-1/07 $25.00 ©2007 IEEE

SIFT features uses the invariable local features to select interest points and then calculate the homography applying these point matches. The images in the same viewpoint but in different directions can be related by Homography. The detail of this algorithm is:
(1). Choose an image as referenced one.
(2). Find the feature matched in the neighboring images.
(3). Calculate the homography H of the two images.
(4). Apply H to warp and project the image 2 to the same coordinate system as the image1, and then process image 2 and stitch them seamlessly.

2. 1. Feature matching

Because larger motions, or matching collections of images where the geometric relationship between them is unknown, it is a key and difficulty to attract features of the images. In our experiment, we use SIFT features and get good results [4].

SIFT features provide a set of features of an object that are not affected by many of the complications experienced in other methods, such as object scaling and rotation. To aid the extraction of these features the SIFT algorithm applies a four stage filtering approach [6]:

(1). Scale-Space Extremum Detection

This stage of the filtering attempts to identify those locations and scales those are identifiable from different views of the same object. This can be efficiently achieved using a "scale space" function. Further it has been shown under reasonable assumptions it must be based on the Gaussian function. The scale space is defined by the function:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \qquad (1)$$

Where * is the convolution operator, G(x, y, σ) is a variable-scale Gaussian and I(x, y) is the input image. Various techniques can then be used to detect stable key point locations in the scale-space. Difference of Gaussians is one such technique, locating scale-space extremum, D(x, y, σ) by computing the difference between two images, one with scale k times the other. D(x, y, σ) is then given by:

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \qquad (2)$$

To detect the local maxima and minima of D(x, y, σ) each point is compared with its 8 neighbors at the same scale, and its 9 neighbors up and down one scale. If this value is the minimum or maximum of all these points, then this point is an extremum.

(2). Key point Localization

This stage attempts to eliminate more points from the list of key points by finding those that have low contrast or are poorly localized on an edge. This is achieved by calculating the Laplacian, value for each key point found in stage 1. The location of extremum, z, is given by:

$$z = -\frac{\partial^2 D^{-1}}{\partial x^2}\frac{\partial D}{\partial x} \qquad (3)$$

If the function value at z is below a threshold value then this point is excluded.

(3). Orientation Assignment

This step aims to assign a consistent orientation to the key points based on local image properties. The key point descriptor can then be represented relative to this orientation, achieving invariance to rotation. The approach taken to find an orientation is firstly to use the key points scale to select the Gaussian smoothed image L, from above and compute gradient magnitude and orientation θ. Secondly, an orientation histogram is formed from gradient orientations of sample points .At last, according to the histogram, orientation to the key points can be assigned.

(4). Key point Descriptor

The local gradient data, used above, is also used to create key point descriptors. The gradient information is rotated to line up with the orientation of the key point and then weighted by a Gaussian. This data is then used to create a set of histograms over a window centered on the key point. Key point descriptor typically uses a set of 16 histograms, aligned in a 4x4 grid, each with 8 orientation bins.

These resulting vectors are know as SIFT keys and are used in a nearest-neighbors approach to identify possible objects in an image. Collections of keys that agree on a possible model are identified, when 3 or more keys agree on the model parameters this model is evident in the image with high probability. Due to the large number of SIFT keys in an image of an object, substantial levels of occlusion are possible while the image is still recognized by this technique

In this experiment, the SIFT features are extracted from a pair of adjacent images. The features are matched using *k-d* tree, and the result is shown in figure 4.

Original images

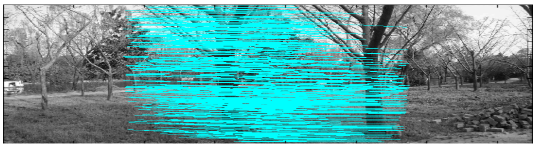

Figure 1: SIFT feature matches extracted from a pair of images

2.2. Homography calculating

As the above described, given two images taken from the same viewpoint but in different directions, the relationship of the two images can be described by a planar projective motion model which is Homography metrics denoted H. This projective transformation warps an image into another one using $x_i' \sim \mathrm{H} x_i$. To calculate H correctly is crucial in the images matching. In our experiment, we use the method of direct linearity transformation (DLT) [3] [5].

H is a matrix of 3×3, so there are 9 elements in total, and we need four pairs matching points at least to calculate H. Assume that given four pairs interest points from image2 to image1, $x_i \leftrightarrow x_i'$ where x_i is from image2 and x_i' from image1. The transformation is $x_i' \sim \mathrm{H} x_i$ where ~ denotes equality up to scale, there exist a nonzero gene between them. So we can describe their relationship as $X_i' \times H\ X_i = 0$, and then we can calculate H. Denoted the j row of H as h^{jT}, then $H\ X_i = \begin{pmatrix} h^{1T} X_i \\ h^{2T} X_i \\ h^{3T} X_i \end{pmatrix}$, We denote $X_i' = (x_i', y_i', w_i')^{\mathrm{T}}$.

So

$$X_i' \times H\ X_i = \begin{pmatrix} y_i' h^{3T} X_i - w_i' h^{2T} X_i \\ w_i' h^{1T} X_i - x_i' h^{3T} X_i \\ x_i' h^{2T} X_i - y_i' h^{1T} X_i \end{pmatrix} \quad (4)$$

Because while j=1，2，3，$h^{jT}\ X_i = X_i^{\ T}\ h^j$,we can get the 9 elements of H by:

$$\begin{bmatrix} 0^T & -w_i' X_i^{\ T} & y_i' X_i^{\ T} \\ w_i' X_i^{\ T} & 0^T & -x_i' X_i^{\ T} \\ -y_i' X_i^{\ T} & x_i' X_i^{\ T} & 0^T \end{bmatrix} \begin{pmatrix} h^1 \\ h^2 \\ h^3 \end{pmatrix} = 0 \quad (5)$$

Where $H = \begin{pmatrix} h^1 \\ h^2 \\ h^3 \end{pmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix}$

Give some remarks on these equations:

(1) $A_i\ h\ = 0$ is a linear equation about h. The elements of matrix Ai are the quadratic polynomial about coordinates of given points.

(2) Although there are three equations in the formula above, but only two equations are linear independent, so when calculating H, the formula can be shortened as shown:

$$\begin{bmatrix} 0^T & -w_i' X_i^{\ T} & y_i' X_i^{\ T} \\ w_i' X_i^{\ T} & 0^T & -x_i' X_i^{\ T} \end{bmatrix} \begin{pmatrix} h^1 \\ h^2 \\ h^3 \end{pmatrix} = 0 \quad (6)$$

(3) This equation also comes into existence for every homogeneous coordinate, so we can let w_i'=1, and the (x_i', y_i') is the coordinate gotten from the image.

2.3. Image stitching

After calculating H, we transform the image 2 into the same coordinate system as image 1 according the formulas introduced in section 2.2 [5]. Because of transformation, lots of information of image can be lost. In this experiment, we record the coordinates of the image 2 after transformation, and then stitch the two images directly. Using this method, complex algorithm and unnecessary calculation are avoided, and higher graphics resolutions can be obtained. The illumination of the two original images is different, but the stitching result is not affected because of the SIFT features' insensitivities to illumination, oriental and scale. The result is shown in figure 2.

Figure2: the image after stitching using the method based on SIFT features

Figure3: the image after stitching using the method based on area [2] (the matching error is marked with red circle)

Some other experiment results based on SIFT features are shown as follow:

Figure4: Original images (with different illumination)

Figure 5: the image after stitching using the method based on SIFT features

Figure6: the image stitched by software

From the result above, a higher graphics resolution is gotten by using our method than by software.

3. Conclusion

Some experiment results have been presented using SIFT feature-based algorithm in this paper. From these results, we can see using SIFT features to transform images allows us stitch them in an accurate way. SIFT feature ensures smooth transformation between images with illumination and orientation differences, and it can also overcome the difficulty of matching in vertical direction. So much higher accuracy and better effect are obtained from the method based on SIFT features in image stitching. At last, because we stitch the two images by using the coordinates of the image after transformation instead of interpolation, unnecessary calculation is avoided.

Because much calculation is needed and the speed of stitching images is a bit slow, the future work is to find an efficient way to reduce the algorithm's complexity.

Acknowledgments

This work is supported by National 973 Key Research Program 2006CB303105, National 973 Key Research Program 2004CB318110 and University Key Research Fund 2004SZ002.

References

[1]. H.Y. Shum and R. Szelishi: Panoramic Image Mosaics, Technical Report, Microsoft Research (1998).

[2]. G. Kang: Virtual Environment Touring based on image, master thesis, Nanjing Institute of Technology (2004).

[3]. R. Hartley，A. Zisserman: Multiple View Geometry in Computer Vision，AnHui University Press (2002).

[4]. H.Y. Shum and R. Szelishi: Construction of Panoramic Image Mosaics with Global and local Alignment, International Journal of Computer Vision 36(2) (2000), 101-130.

[5]. Y. Li, L. Ma: A Fast and Robust Image Stitching Algorithm, in proceeding of the 6th World Congress on Intelligent Control and Automation (2006).

[6]. Hui Zhao, Matching Technology based on SIFT Features, Shan Dong University, Institute of Information Science and Engineering, (2006).

[7]. D. Lowe. Distinctive image features from scale-invariant keypoints. Int. Journal of Computer Vision, 60(2):91–110, (2004).