A Method for Generation of Panoramic View based on Images Acquired by a Moving Camera

Prajkta Sangle
Dept. of Electronics and Telecommunication, Cummins College of Engineering for Women, India.

Krishnan Kutty
Centre for Research in Engineering Sciences and Technology, KPIT Cummins Infosystems Limited, India.

Anita Patil
Dept. of Electronics and Telecommunication, Cummins College of Engineering for Women, India

ABSTRACT

Panoramic photo stitching is the process of combining multiple photographic images with overlapping fields of view to produce a panorama. The process to generate a panoramic view can be divided into three main components - image acquisition, image registration, and blending. In this paper, a robust algorithm called Scale Invariant Feature Transform (SIFT) used to extract the features from the images and matching them which is a part of image registration. SIFT features are invariant to rotation, translation, image scaling and partially invariant to 3D viewpoint, illumination changes and image noise. Image transformation is estimated using homography. Image blending technique is used to blend the images together to get a panoramic view. Main applications of panoramic view include creating virtual environment for virtual reality, modeling the 3D environment using images acquired from the real world.

General Terms

Image processing, computer vision

Keywords

Panoramic view, SIFT, homography, image blending

1. INTRODUCTION

Image stitching or photo stitching is the process of combining multiple photographic images with overlapping fields of view to produce a panorama. The automatic construction of large, high-resolution image mosaics is an active area of research in the fields of photogrammetry, computer vision, image processing, and computer graphics [1]. In computer vision, image mosaics are part of a larger recent trend, namely the study of visual scene representations. In graphics, image mosaics play an important role in the field of image based rendering, which aims to render photorealistic views from collections of real world images. For applications such as virtual travel and architectural walkthroughs, it is desirable to have complete panoramas, i.e., mosaics which cover the whole viewing sphere and hence allow the user to look in any direction. Image alignment algorithms can discover the correspondence relationships among images with varying degrees of overlap. Panoramic view generation algorithms take the alignment estimates produced by registration algorithms and blend the images in a seamless manner.

1.1 Stages for Generation of Panoramic View

There are three main stages in the process of generation of panoramic view, viz. image acquisition, image registration and image blending. Preprocessing is sometimes needed for the images which are obtained by a wide angle lens. The type of distortion introduced may be radial or pincushion depending on the type of lens and application. Also, postprocessing is the means of getting seamless panorama which is sometimes included in the blending algorithm. Therefore the stages are described here.

1.1.1 Image Acquisition

Typically for panoramic imaging, images can be acquired by three methods namely translating a camera parallel to the scene or rotating a camera about its vertical axis keeping optical centre fixed or by a handheld camera. Each image in the series acquired for panoramic image stitching partially overlaps the previous and the following images. Images acquired by translating a camera does not give 3D feel to the panoramic image and is generally not preferred. Rotation of camera provides this 3D feel [5].

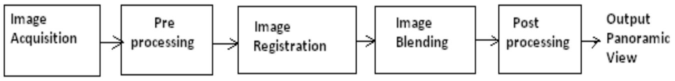


Figure 1 Block diagram of panoramic image stitching

Due to the simplicity of the set-up and the operations, rotation method is preferred in acquiring images for the generation of panoramic view. Use of handheld camera is the most challenging case to handle because of varying overlaps and transformation in the images.

1.1.2 Image Registration

To form a larger image with a set of overlapping images, it is necessary to align the images. The process of image registration aims to find the transformations to align two or more overlapping images such that the projection from the view point through any position in the aligned images into the 3D world is unique [6].

Direct or feature-based image alignment methods may be used. Direct alignment methods search for image orientations that minimize the sum of absolute differences between overlapping pixels. When using direct alignment methods one might first calibrate their images to get better results. Feature-based methods determine proper image orientations by identifying features that appear in multiple images and overlapping them.

Features consist of significant parts of image such as corners, points, lines, edges. Only neighboring images are searched for matching features. Since there are smaller group of features for matching the results of the search is thus more accurate and execution of the comparison is faster. When we get the matching features, the image transformation is estimated so that the resultant images are geometrically aligned or registered.

1.1.3 Image Blending

Once the overlapping images have been registered, they need to be merged together to form a single panoramic image. Image blending or merging is the process of adjusting the values of pixels in two registered images, such that when the images are joined, the transition from one image to the next is invisible. The process of image merging is performed to make the transition between adjacent images visually undetectable. At the same time, the merged images should preserve the quality of the input images as much as possible [12].

Due to various reasons, including the lighting condition, the geometry of the camera set-up and other reasons, the overlapping regions of adjacent images are almost never the same. Therefore, removing part of the overlapping regions in adjacent images and concatenating the trimmed images often produce images with distinctive seams. A seam is the artificial edge produced by the intensity differences of pixels immediately next to where the images are joined.

2. LITERATURE REVIEW

The methods for automatic image matching fall broadly into two categories as direct and feature based matching. Direct methods minimize the sum of absolute differences between overlapping pixels or use any other cost functions available. These methods are computationally complex as they compare each pixel window to other and are not invariant to image scale and rotation.

Feature based methods begin by establishing correspondences between points, lines, edges, corners or other geometric entities. For example, an approach proposed by Harris and Stephens in 1988 used Harris corner to detect the features and use a normalized cross-correlation of intensity values to match them [8]. Harris corner is not invariant to scale changes and cross-correlation is not invariant to rotation. Hence these methods are not suitable for robust feature matching [15].

The work of Schmid and Mohr (1997) reveals that image recognition problems can be solved by invariant local feature matching in which a feature was matched against a large database of images [10]. They used Harris corners to select interest points, but rather than matching with a correlation window, they used a rotationally invariant descriptor of the local image region. This allowed features to be matched under arbitrary orientation change between the two images. Also they used a multiscale approach to make feature detection scale invariant but there is no mention about the invariance to illumination changes and change in affinity of the features.

David Lowe proposed a distinctive descriptor called Scale Invariant Feature Transform (SIFT) in 1999 for object recognition [7] and extended the proposed work in 2004 [2] making the descriptor more robust and invariant to rotation, translation, scale and partially invariant to changes in 3D viewpoint and illumination. In the survey paper proposed by Mikolajczyk and Schmid in 2005, it is concluded that SIFT and its extension gives most accurate results for feature detection [10]. In the paper proposed by Lowe and Brown this method has successfully used for panoramic stitching [1].

3. PROPOSED METHODOLOGY

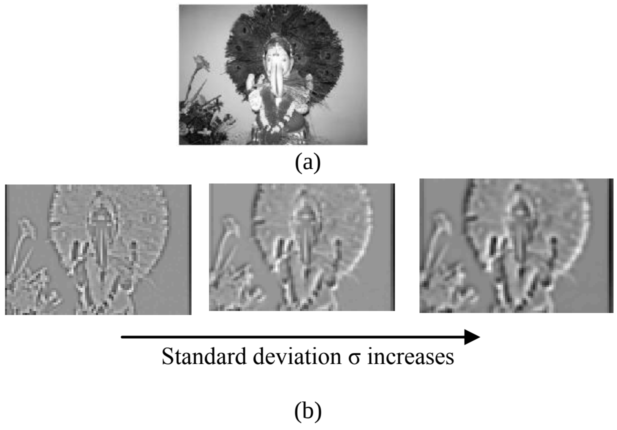
In this paper we describe a feature based approach to create a panoramic view. The images are acquired by an ordinary lens with the help of digital camera. So there is no distortion in the images. We directly start with registration of acquired images. Image registration consists of extracting the features, matching them and estimating transformation. Firstly, use of invariant features enables reliable matching of panoramic image sequences despite rotation, zoom and illumination change in the input images. Secondly, we can discover the matching relationship between the images which is known as transformation estimation. This estimation is done by homography estimation. Thirdly, high-quality results are generated using blending to render output panorama.

3.1 Feature Detection

Scale Invariant Feature Transform (SIFT) is an approach for detecting and extracting features that are reasonably invariant to changes in illumination, image noise, rotation, scaling, and small changes in viewpoint as proposed by David Lowe [2]. The features are highly distinctive, in the sense that a single feature can be correctly matched with high probability against a large database of features from many images.

The main feature of SIFT which makes its performance outstanding among already existing image registration algorithms is that SIFT features share a number of properties in common with the responses of neurons in inferior temporal (IT) cortex in primate vision [4].

The first step in the panoramic recognition algorithm is to extract and match SIFT features between all of the images. SIFT features are located at scale-space extrema of a difference of Gaussian function [3].


(a)

(b)

Figure 2 (a) Original image (b) Difference of Gaussian Scale space with values of sigma varying from 1.6 to 3.2

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \qquad (1)$$

G(x, y, σ): Image scaled by scaling factor σ
σ : Standard deviation of Gaussian kernel

Keypoints are selected based on measures of their stability [13]. Keypoints with low contrast and the keypoints located along the edges are rejected. At each feature location, a characteristic scale and orientation is established. This gives a similarity-invariant frame in which to make measurements. The invariant descriptor is actually computed by accumulating local gradients in orientation histograms. This allows edges to shift slightly without altering the descriptor vector.

Illumination invariance is achieved by using gradients and normalizing the descriptor vector [14]. Since SIFT features are invariant under rotation and scale changes, algorithm can handle images with varying orientation and zoom. This would not be possible using traditional feature matching techniques such as correlation of image patches around Harris corners. Ordinary (translational) correlation is not invariant under rotation, and Harris corners are not invariant to changes in scale.

3.2 *Feature Matching*

At this stage the objective is to find matches between successive images. When the sift descriptors are formed, they are compared between the two images, and the descriptors with minimum Euclidian distance between them are said to have matched. If the ratio of 1st match to 2nd match is less than 0.8, then the match is selected otherwise discarded [2].

3.3 Homography Estimation

In homogeneous coordinates, all geometric transformations can be written as matrix multiplication. Let the points in the first image be (x1 y1), …, (x4, y4), and their corresponding points in the second image be (u1, v1), … , (u4, v4). Then for each corresponding pair of points, we can obtain equation as [11]:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \qquad (2)$$

Although there are nine unknowns a, b, c, d, e, f, g, h and i in the homography matrix, only eight of them need to be calculated because we are working in homogeneous coordinates. It is customary to let i = 1 and then seek to determine the other unknowns. Rewriting all the equations in terms of the unknowns a, b, c, d, e, f, g, h we get an 8x8 system as given by equation (3).

$$\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -u_1x_1 & -u_1y_1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -u_2x_2 & -u_2y_2 \\ x_3 & y_3 & 1 & 0 & 0 & 0 & -u_3x_3 & -u_3y_3 \\ x_4 & y_4 & 1 & 0 & 0 & 0 & -u_4x_4 & -u_4y_4 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -v_1x_1 & -v_1y_1 \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -v_2x_2 & -v_2y_2 \\ 0 & 0 & 0 & x_3 & y_3 & 1 & -v_3x_3 & -v_3y_3 \\ 0 & 0 & 0 & x_4 & y_4 & 1 & -v_4x_4 & -v_4y_4 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} \qquad (3)$$

So, 4 correspondences are necessary to get the transformed image. These correspondences are selected such that these are located as far as possible from each other.

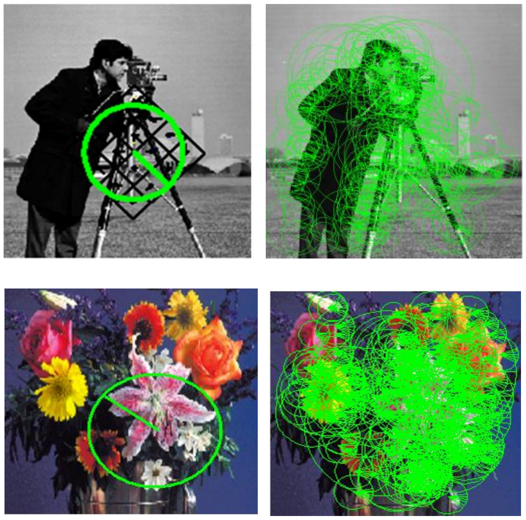
3.4. Image Blending

Image blending is a common practice in the generation of panoramic images and applications such as object insertion, super resolution and texture synthesis. The images should be stitched to generate a mosaic image. A simple pasting of images produce visible artificial edges in the seam between the images, due to differences in camera gain, scene illumination or geometrical misalignments.

If we have a set of images, the reference image is chosen. Second image in the set is transformed to get aligned with the reference image. This gives us a blended image of first two. Now this is treated as the reference and third image is transformed according to new reference. In this way all images are blended and a final panoramic image is obtained.

4. RESULTS

Results are obtained by making slight changes in the original paper by David Lowe. Here, the size of image is not doubled, but used as it is for constructing a Gaussian pyramid. The centre of the circle denotes keypoint detected. Radius of the circle is equal to six times the scale at which keypoint is detected. 4x4 image patch surrounding the keypoint shows 16 pixels and their corresponding orientations. SIFT does not use color information present in the image so first the color images are converted into grayscale and then SIFT is applied. However, we can show SIFT features detected on color images, because conversion of color image to grayscale doesn’t change spatial information of the image.



(a) **(b)**

Figure 3 (a) Keypoint detected and its dominant orientation

(b) all SIFT features detected in 250x250 image

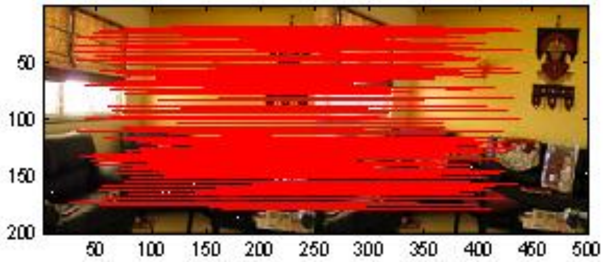
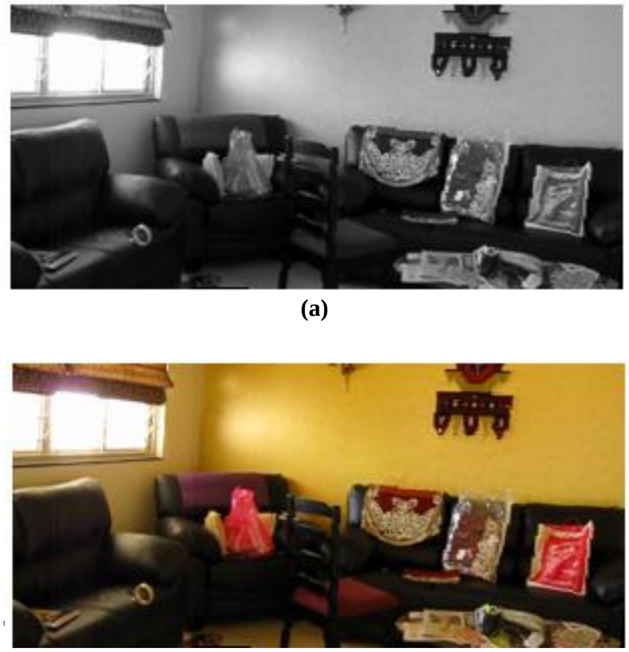


Figure 4: Features detected on a real-time image by Scale Invariant Feature Transform (SIFT) and matched by using nearest neighbor algorithm



(a)

(b)

Figure 5: Result of blending multiple images (a) Grayscale image (b) Color Image

5. CONCUSION

We conclude that Scale invariant feature transform can be used to extract features from the images. The features are invariant to rotation, scale and partially invariant to illumination change and image noise. These features can be efficiently used to match the corresponding features from overlapping images. To discard the spurious matches ratio of 1st nearest neighbor to 2nd nearest neighbor is used which gives accurate results. By using these features, a transformation is estimated and the images are blended to get a panoramic view.

6. REFERENCES

[1] M. Brown, D. Lowe, "Automatic Panoramic Image Stitching using Invariant Features", International Journal of Computer Vision 74(1), 59–73, 2007, Springer Science + Business Media, LLC. Manufactured in the United States

[2] David Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, 2004.

[3] Lindeberg, T. 1993. "Detecting salient blob-like image structures and their scales with a scale-space primal sketch: a method for focus-of-attention" International Journal of Computer Vision, 11(3): 283-318.

[4] Ito, Minami, Hiroshi Tamura, Ichiro Fujita, and Keiji Tanaka, "Size and position invariance of neuronal responses in monkey inferotemporal cortex," Journal of Neurophysiology, 73**,** 1 (1995), pp. 218–226.

[5] C. Chen and R. Klette, "An image stitcher and its application in panoramic movie making." Proc.DICTA'97, Dec.1997, pp.101-106.

[6] Szeliski, R. 2004. "Image alignment and stitching: A tutorial." Technical Report MSR-TR-2004-92, Microsoft Research.

[7] David Lowe, "Object recognition from local scale-invariant features" In International Conference on Computer Vision, Corfu, Greece, pp. 1150-1157

[8] Harris and M.J. Stephens, "A combined corner and edge detector" In Alvey Vision Conference, pages 147–152, 1988.

[9] Mikolajczyk and Schmid, 2005"A performance evaluation of local descriptors", In European Conference on Computer Vision (ECCV), Copenhagen, Denmark

[10] Schmid, C., and Mohr, R. 1997, "Local grayvalue invariants for image retrieval" IEEE Trans. on Pattern Analysis and Machine Intelligence, 19(5):530-534.

[11] Xi Shao, Changsheng Xu, Joo Hwee Lim, "Image Mosaics Base on Homogeneous Coordinates" Institute for Infocomm Research.

[12] Anat Levin, Assaf Zomet, Shmuel Peleg, and Yair Weiss, "Seamless Image Stitching in the Gradient Domain", research supported (in part) by the EU under the Presence Initiative through contract IST-2001-39184, Benego.

[13] Yu Meng and Bernard Tiddeman, "Implementing the Scale Invariant Feature Transform (SIFT) Method", Department of Computer Science University of St. Andrews

[14] Andrea Vedaldi, "An implementation of SIFT detector and descriptor", University of California at Los Angeles

[15] Konstantinos G. Derpanis, "The Harris Corner Detector".