http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 10 (10): 1977-1982, 2011
ISSN 1812-5638 / DOI: 10.3923/itj.2011.1977.1982

Automatic Panorama Creation using Multi-row Images

Baosen Song, Yongqing Fu and Jinlin Wang
Room 508, No. 21 Building, College of information and Communication Engineering,
Harbin Engineering University, No. 145, Nantong Street, Harbin,
Heilongjiang Province, Zip Code 150001, China

Abstract: Restricting by cameras' angles of view of the surveillance systems, the panoramic digital images can not be obtained directly from photographing. To enlarge the horizontal and vertical angles of view for an image, an automatic panorama creating method was proposed, the key techniques of the method including camera calibration, image feature extraction, image registration, bundle adjustment, photometric optimization, image fusion and output panorama projection were also explored. An automatic panoramic image mosaicing was realized by Microsoft Visual Studio 2008 using C++. Experiments show that the software can stitch together multi-row images automatically into a panoramic image. In addition, it also reduces visual distortions, making the panoramic view clearer and its luminance more homogeneous and natural.

Key words: Panorama stitching, multi-row image mosaic, photometric optimization, image localization

INTRODUCTION

Automatic panoramic image mosaicing attracts a great deal of attention addresses the Angle-of-View (AoV) limitations of an imaging device that cannot capture a wide-angle image at once (Brown and Lowe, 2007). Its potential applications are extensive, ranging from medical image synthesis, remote sensing image processing, roaming virtual environments and computer vision, among many others. Therefore, an in-depth study of automatic high-quality panoramic image mosaicing is necessary and meaningful.

The former method of creating panorama used Harris corner feature to obtain registration of images (Li and Du, 2010) but Harris corner feature is severely disturbed by scale change of images. To solve this problem, the Scale Invariant Feature Transform is used (Hua *et al*., 2010) but the method of Hua can stitch only two images, because the method didn't use the bundle adjustment technology to cancel the error of multi-image registration. A better method (Xiong and Pulli, 2010) of panorama creating on mobile can correct color difference between images and use little memory. But the mentioned methods have two common defects: the first is that the evidence brightness difference between images still exists after stitched; second is that they attend only to the horizontal AoV and neglect the vertical AoV. Although, the method of Brown can stitch together more than two rows of images into a panoramic image, it causes substantial visual distortions in vertical direction.

In order to solve the first problem, photometric optimization technology has been used to uniform the brightness and color difference of the panorama. To enlarge the vertical angles of view for the panorama, it must be stitched together by using more than two rows of images. The improved cylindrical projection algorithm was proposed to reduce visual distortions, especially the vertical distortions.

ALGORITHMS

The proposed approach, outlined below, constitutes a feature-based panorama mosaicing technology. First, the input images are pre-processed. Second, the Edge-group Scale Invariant Feature Transform (EG-SIFT) (Fu *et al*., 2010) algorithm is used to extract image features. Third, the features between images are matched to realize image registration. Fourth, all images are fitted to their positions in the final panoramic view. Fifth, the projection matrix of each image is optimized by using a bundle adjustment technology. Sixth, the images are processed with photometric adjustments. Finally, the images are blended with a multi-resolution fusion algorithm to obtain the panoramic image. In the rest of this section, the main algorithms of proposed method are described.

Image pre-processing: The first step in image pre-processing is correcting the lens parameters of the camera (Arif *et al*., 2002; Pu *et al*., 2011). Subsequently, the image

Corresponding Author: Baosen Song, Room 508, No. 21 Building, College of Information and Communication Engineering, Harbin Engineering University, No. 145, Nantong Street, Harbin, Heilongjiang Province, Zip Code 150001, China

is transformed according to the corrected parameters to remove image distortions caused by the lens.

Lens distortion parameters: For an ideal rectilinear camera lens, perfect results can be achieved by mapping pixels in the image to the tangent plane. However, real lenses deviate from this perfect tangent plane projection. The deviations lead to points in the scene being fixed away from where they would have fallen. Fortunately, almost all deviations occur radially, towards or away from some common center and the amount of deviation is almost the same at a given radius around that center. Hence, a model that corrects for this deviation based on the radius gives satisfactory results.

The lens distortion parameters a, b, c and d correspond to the parameters of a third degree polynomial describing the radial lens distortion:

$$r_{src} = (ar_{dest}^3 + br_{dest}^2 + cr_{dest} + d)r_{dest} \quad (1)$$

where, r_{dest} and r_{src} are the normalized radii of image pixels. Here, the term normalized indicates that the largest circle fitting completely into an image has a radius equal to 1. The fourth parameter d controls the resulting image size and is calculated as follows:

$$d = 1-(a+b+c)$$

For a perfect lens, a = b = c = 0 and d = 1.0. This resolves into $r_{dest} = r_{src}$. This polynomial approach provides a good approximation to the actual behavior of a given lens.

Lens shift parameters: Occasionally, a lens and image sensor may be incorrectly aligned so that the optical axis fails to fall on the image center.

The lens shift parameters e (horizontal shift) and g (vertical shift) compensate for this problem. Their values which are in pixel units, determine the extent to which how far the center for radial correction is shifted outside the geometrical center of the image.

Field of view: The term field of view is often used to refer to the AoV. The AoV of a photograph or camera is a measure of the proportion of a scene included in the image. Hereafter, reference will be made only to the Field of View (FoV). The FoV, along with the type of lens projection, determines the image perspective distortion.

For a rectilinear lens, the FoV can be calculated from the focal length as follows:

$$FoV = 2 * \text{atan}\left(\frac{\text{size}}{2 * \text{FocalLength}}\right) \quad (2)$$

where, size is either the width or height for the respective FoV, atan (x) is arctangent function of x.

Local registration: The aim of local registration is to achieve the correct projection (Ravichandran and Ravindran, 2007; He *et al*., 2011). Relation for every pair of images with overlapping areas. It includes two steps: image feature extraction and image feature matching.

Image feature extraction: In this study, EG-SIFT is used to extract image features. EG-SIFT consists of five major stages: (1) edge group detection; (2) scale-space peak detection according to edge group; (3) accurate localization of keypoints; (4) assignment of keypoint orientation; (5) construction of descriptors.

After these steps, feature points with the following characteristics can be obtained: (1) valued for their location at the edge of an image; (2) invariant to shifts, rotations and scaling of the image, as well as changes in illumination and (3) partially invariant to noise, changes in view and affine image distortions.

Image feature matching: The best match for each feature point is obtained by identifying its nearest neighbor among the feature points of input images (Feng-Dong *et al*., 2009). The nearest neighbor is defined as the feature point with the minimum Euclidean distance of the EG-SIFT vector.

However, many features of an image will not have a correct match in other images because they arise from non-overlapping areas. Therefore, it is necessary to discard such no-match features. A more effective method is to compare the distance of the closest neighbor to that of the second-closest neighbor. This method performs well because in correct matches, the nearest neighbor is significantly closer than the nearest incorrect match to achieve reliable matching.

In this study, the Best-bin-first algorithm is used to identify the nearest neighbors of feature points, because it is suitable for the high dimensionality of the EG-SIFT vector.

Global adjustment: The aim of this step is to identify the spatial relationship of the images, obtain the best projection transform matrix of the image map to the final panoramic plane and to some extent solve luminance and chromatic aberrations.

Image location: It is more difficult to locate images in a multi-row mosaic than in a single-row mosaic. In a multi-row mosaic, each image is equivalent to a point on a two-dimensional plane. Thus, each image is expressed

as I_{pq}, where p and q are the column row indices, respectively. The relationship expressed by I_{pq} (hu, hd, vl, vr) $\cap$ other denotes that the four sides of image i_{pq} overlap with the corresponding sides of the other images, where hu, hd, vl and vr denote the top, bottom, left and right side of the image. During the locating process, the first priority is to find the top-left image, I_{00}. Once this is accomplished, all other images can be easily located. The images in the top-most row, I_{p0}, are obtained by the following formula:

$$I \in \overline{I_{pq}(hu) \cap other} \tag{3}$$

If all images in the top-most row I_{p0} sum to 360° horizontally, then I_{00} is obtained by the following:

$$I \in I_{p0}(hd) \cap other \tag{4}$$

If I_{p0} sum to less than 360°, then I_{00} is obtained by the following:

$$I \in I_{p0}(vl) \overline{\cap} other \tag{5}$$

Once I_{00} is found, the first image of the second row, I_{01}, is found easily because its top side overlaps with I_{00}. The second-row images can then be located. The same process is repeated until all images are located.

Bundle adjustment: In the local registration process, the projection transform matrix is obtained for each pair of images. As the ultimate aim is to map all images to the panoramic plane, the optimal transform matrix for each image map to the panoramic plane must be found. In this study, bundle adjustment (Liu *et al*., 2008) technique is used to find the optimal matrix.

During the adjustment process, the algorithm reads in the images in order and optimizes the projection transform matrix of each image to the panoramic plane. The aim of the optimization method is to calculate the distance from a feature point in image I_i to its corresponding point in image I_j, by minimizing the distance value to adjust the transform matrix of each image map to the panoramic plane to achieve optimization. The adjustment process is shown in Fig. 1.

For pair-wise features (x_i, x_j), x_i is mapped to the panoramic plane and then to the image containing x_j as $x`_i$. The distance is obtained as follows:

$$r_{ij} = \left\| x_j - H_j^{-1} H_i x_i \right\| \tag{6}$$

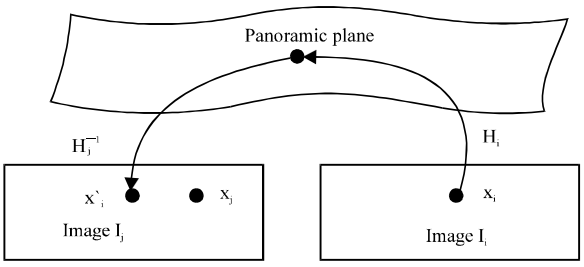

Fig. 1: Theory of bundle adjustment

where, r_{ij} is the difference in distance and H is the projection transform matrix. The sums of the differences in distance are calculated for all pair-wise feature points that belong to all images. The target function for optimization is thus obtained as follows:

$$err = \sum_{i=1}^{n} \sum_{j \in L(i)} \sum_{\lambda \in F(i,j)} Q(r_{ij}^{\lambda})^2 \tag{7}$$

where, L (i) denotes the set of all images overlapped with image I_i, F (i, j) denotes the set of all pair-wise features between image I_i and image I_i and r^{λ}_{ij} denotes the distance r of pair-wise feature λ. The transform matrix H_i is then adjusted by updating the follow formula:

$$H_i = H`_{ik} * H_i \tag{8}$$

where, $H`_{ik}$ is the updating quantity of the transform matrix between image I_i and I_k.

The solution to Eq. 7 is given by the L-M algorithm (Lale *et al*., 2002).

Photometric optimization: In order to achieve better image quality, the image must be processed by a photometric adjustment technology after bundle adjustment (Doutre and Nasiopoulos, 2009).

The image irradiance E incident on the image plane is then integrated at each pixel on the imaging sensor with an exposure time t, resulting in a measurement of energy Q. After scaling the measured energy Q with the gain factor s, the resulting value is subjected to a camera response function f. The grey value B measured by the camera is thus given by:

$$B = f(eML) \tag{9}$$

where, the effective exposure is given by:

$$e = \pi\, st/k^2$$

where, L is the scene radiance and M is a function describing the spatially varying attenuation caused by the

optical system. The following radial polynomial is then used:

$$M = \beta_1\gamma^6+\beta_2\gamma^4+\beta_3\gamma^2+1$$

where, γ is the Euclidean distance of a point x in the image plane from the center of the vignetting. Except for exotic cameras, f is a monotonically increasing function and an inverse camera response function $g = f^{-1}$ exists. The radiance of a scene point can then be reconstructed from the grey value B by solving Eq. 9 for L:

$$L = \frac{g(B)}{eM} \quad (10)$$

For color cameras, assuming that the exposure e of a channel i is scaled by a white balance factor w_i and the same response function f is applied to each channel, B is obtained as follows:

$$B = f(w_i eML) \quad (11)$$

Suppose that two images of a static scene have been captured with different exposure times or camera orientations. For the same point, two different grey values B_1 and B_2 are measured owing to the different exposure times or a spatially varying vignetting term M. Assuming the same values for scene point radiance L:

$$\frac{g(B_1)}{e_1M(x_1)} = \frac{g(B_2)}{e_2M(x_2)} \quad (12)$$

By solving Eq. 12 for B_1, the grey value transfer function τ is obtained as follows:

$$B_1 = \tau(B_2) = f\left(g(B_2)\frac{e_1M(x_1)}{e_2M(x_2)}\right) \quad (13)$$

Assuming a discrete response curve f defined for grey levels between 0 and 255, a monotonous response curve can be enforced by using the following equation:

$$e_m = \sum_{i=1}^{255}(\min(f(i)-f(i-1),0))^2 \quad (14)$$

Finally, a set of N pairs of grey values B_1 and B_2 are obtained. These parameters can be estimated by minimizing the error term below:

$$e = Ne_m + \sum_i^N d(B_{i1} - \tau B_{i2}) + d(B_{i2} - \tau^{-1}B_{i1}) \quad (15)$$

where, d (x) is the Euclidean distance. The L-M algorithm is then used to solve Eq. 15.

Output image projection: When the full or partial 3D scene is mapped onto a two-dimensional (2D) print or screen (Her *et al*., 2011), a stretching distortion usually occurs when a viewing sphere is projected onto the cylindrical image surface in an oblique direction. In contrast to the conventional cylindrical projection, an adjustable cylindrical surface is used in this study to match the viewing direction with the equator of the cylindrical surface. In addition, in order to find the balance point between the stretching distortion and the bending of straight lines, the curvature of the cylindrical surface is adjusted according to the object of interest in the image. This requires that the sphere is rotated and that the radii of both the sphere and cylinder are adjusted. In order to reduce the large stretching distortion, the cylinder must be tilted and rotated so that the projection vector $\vec{p}$ becomes close to the equator vector $\vec{z}$. This is actually the same as rotating the viewing sphere. Hence, the viewing sphere is first rotated before warping, according to Rodriquez's formula:

$$R(\hat{n},\theta) = I + \sin\theta[\hat{n}]_\times + (1-\cos\theta)[\hat{n}]_\times^2 \quad (16)$$

where, R is the rotation matrix, $\hat{n}$ is the rotation axis and θ is the rotation angle (Yang *et al*., 2010).

After rotating the viewing sphere, the curvature of the cylindrical surface is adjusted. In the case of perspective projection, the curvature of the surface is zero. However, the cylindrical projection has a surface with a radius of δ, i.e., a plane with a curvature of $1/\delta$. Hence, the curvature is adjusted between 0 and $1/\delta$, depending on the main features in the corresponding position.

Image fusion: Wavelet multi-resolution image fusion technology (Sasikala and Kumaravel, 2007) is used in this study, given that Mallat presented a fast wavelet algorithm of the discrete dyadic wavelet transform. For a 2D image signal f (x, y) under scale j, the wavelet decomposition algorithm can be expressed as follows:

$$\begin{aligned} A_{j+1} &= H_c H_r A_j \quad B_{j+1} = G_c H_r A_j \\ C_{j+1} &= H_c G_r A_j \quad D_{j+1} = G_c G_r A_j \\ A_j &= H_r^* H_c^* A_{j+1} + H_r^* G_c^* B_{j+1} + G_r^* H_c^* C_{j+1} + G_r^* G_c^* D_{j+1} \end{aligned} \quad (17)$$

where, H is a low-pass filter, G is a high-pass filter, A is the set of approximate coefficients, B is the set of horizontal detail coefficients, C is the set of vertical detail coefficients and D is the set of diagonal detail coefficients. In addition, H* and G* correspond to the conjugate transposed matrices of H and G, respectively where, A_{j+1}, B_{j+1}, C_{j+1} and D_{j+1} correspond to the low-frequency, horizontal high-frequency, vertical high-frequency and diagonal high-frequency components, respectively. The corresponding 2D reconstruction equation for A_j is also shown above.

IMPLEMENTATION OF APIMS

In this study, the automatic panoramic image mosaicing system (APIMS) is realized in Microsoft Visual Studio 2008 by using the C++ code. A flowchart of APIMS is shown in Fig. 2.

The detailed descriptions for each step are as follows:

Step 1: Multi-select wxFileDialog is used to load n input images. All GUIs of the APIMS are designed by using a cross-platform GUI and a tools library called wxWidgets 2.8.11, of which wxFileDialog is one class

Step 2: Every image is checked for whether it contains a valid parameter. If not, valid parameters must be inputted by manual entry or by loading calibrated lens data. Otherwise, the lens should be calibrated

Step 3: The parameters a, b, c, d, e, g and HFoV are calculated using the valid parameters and the EXIF information of every input image. The image is then corrected by using a, b, c, d, e and g

Step 4: A number of feature points with a 128-dimensional vector are obtained by extracting the EG-SIFT feature of every image

Step 5: The BBF algorithm is first used during image registration to find the minimum Euclidean distance to match two feature points (the feature matching step). In the second step, poorly matching feature points are removed with the RANSAC algorithm

Step 6: The pair-wise features resulting from step 5) are used to calculate all transform matrices (H) that map points from one image projecting onto a neighbor image

Step 7: During the process of locating n input images, q images that have been located are labeled. Subsequently, the location of the remaining n-q images are found to reduce the time spent on this process

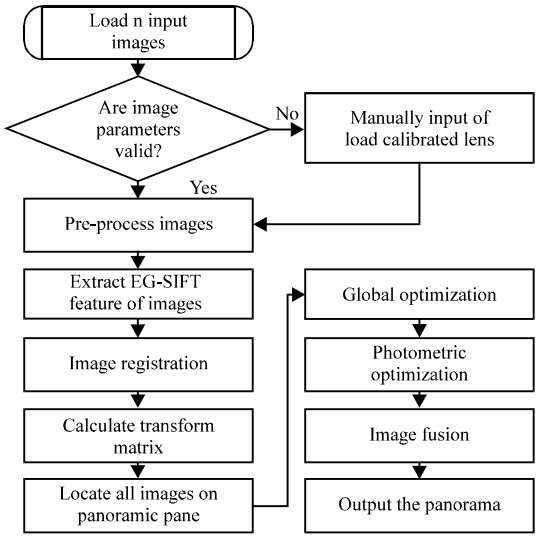

Fig. 2: Block diagram of APIMS software

Step 8: Bundle adjustment is used to optimize the transform matrix H from every image projecting to the panoramic plane by minimizing Eq. 7 and modifying H by Eq. 8

Step 9: By minimizing Eq. 15 with the L-M algorithm to optimize the photometric parameters, the image to the panoramic plane is transformed according to the optimized parameters

Step 10: Wavelet multi-resolution image fusion technology is used to remove the stitch slots of the panoramic images

Step 11: Finally, as described above, the panoramic image is outputted by using the improved cylindrical projection

EXPERIMENTAL RESULT

In order to demonstrate the advantages of presented APIMS software, a scene of the Orekhov plaza of Harbin Engineering University was obtained with a digital camera (Canon IXUS 130). The group photos are stitched together into a panorama by the method of Brown and Lowe (2007) and proposed method. Figure 3a is the result panorama created by Autostitch, Autostitch is software designed by Brown according to his panorama creating method. Figure 3c is the result panorama created by presented APIMS software.

From Fig. 3b, the distortion of the high building is more severe than that of Fig. 3c, so APIMS can reduce the vertical distortion of the panorama in some degree. In Fig. 3c, the brightness and color of the

Fig. 3(a-c): Original images and resulting panoramic images

panorama is more homogeneous than that of Fig. 3b, so the presented photometric optimization technology is effective to remove the photometric difference between images.

CONCLUSION

In this study, C++ code is used to realize a method for automatically creating panorama images from multi-row images. APIMS outputs high-quality panoramic images and outperforms other procedures in the following two aspects: 1) reducing visual distortions, especially in the vertical direction and 2) making the panoramic image clearer and its luminance more homogeneous and natural.

REFERENCES

Arif, M., H. Xinhan and W. Min, 2002. Stratified approach to 3D reconstruction. Inform. Technol. J., 1: 75-79.

Brown, M. and D.G. Lowe, 2007. Automatic panoramic image stitching using invariant features. Int. J. Comput. Vision, 74: 59-73.

Doutre, C. and P. Nasiopoulos, 2009. Fast vignetting correction and color matching for panoramic image stitching. Proceedings of the 16th IEEE International Conference on Image Processing, Nov. 7-10, Cairo, Egypt, pp: 709-712.

Feng-Dong, C., B.R. Hong and G.D. Liu, 2009. Planar displacement detection with point feature matching. Inform. Technol. J., 8: 383-387.

Fu, Y., B. Song and J. Wu, 2010. An improved scale invariant feature transform algorithm. J. Harbin Eng. Univ., 31: 632-636.

He, C., F. Lang and H. Li, 2011. Medical image registration using cascaded pulse coupled neural networks. Inform. Technol. J., 10: 1733-1739.

Her, M.G., C.C. Peng, K.S. Hsu and T.H. Liu, 2011. Development of a new single beam omnidirectional projector and its application in tele-robotic system. Inform. Technol. J., 10: 816-824.

Hua, Z., Y. Li and J. Li, 2010. Image stitch algorithm based on SIFT and MVSC. Proceedings of the FSKD 7th International Conference, Aug. 10-12, Yantai, China, pp: 2628-2632.

Lale, O., T. Tulay and K. Kevser, 2002. Comparison of neural algorithms for function approximation. J. Applied Sci., 2: 288-294.

Li, J. and J. Du, 2010. Study on panoramic image stitching algorithm. Proceedings of the 2nd Pacific-Asia Conference on Circuits, Communications and System, Aug. 1-2, Beijing, China, pp: 417-420.

Liu, S., J. Sun and J. Dang, 2008. A linear resection-intersection bundle adjustment method. Inform. Technol. J., 7: 220-223.

Pu, Y., S. Lee and C. Kuo, 2011. Calibration of the camera used in a questionnaire input system by computer vision. Inform. Technol. J., 10: 1717-1724.

Ravichandran, C.G. and G. Ravindran, 2007. New fully automatic fast registration method for 2D computed tomography images. Inform. Technol. J., 6: 761-765.

Sasikala, M. and N. Kumaravel, 2007. A comparative analysis of feature based image fusion methods. Inform. Technol. J., 6: 1224-1230.

Xiong, Y. and K. Pulli, 2010. Fast panorama stitching for high-quality panoramic images on mobile phones. IEEE Trans. Consum. Electron., 56: 298-306.

Yang, C., H. Mao, G. Abousleman and J. Si, 2010. Correction of projective distortion in long-image-sequence mosaics without prior information. Proc. SPIE, Vol. 7668, 10.1117/12.849854.