Multispectral Panoramic Mosaicing

Udhav Bhosle, Sumantra Dutta Roy and Subhasis Chaudhuri
Department of Electrical Engineering
Indian Institute of Technology Bombay, Powai, Mumbai -400 076
{udhav,sumantra,sc}@ee.iitb.ac.in

Abstract

Image mosaicing overcomes the limitations of a camera's limited field of view by aligning and pasting frames in video sequences. We introduce the concept of multispectral mosaicing. The information from IR and visual images is directly fused at the pixel level so that subsequent operations can be carried out using the fused image instead of the individual sensor images. We develop a geometric relationship between a visual band and IR band panoramic mosaic. Our system uses a fast algorithm for automatic construction of panoramic mosaic. We show results of inter-band mosaic superposition in support of the proposed strategies.

1. Introduction

In mosaicing, images are captured as a camera moves, registered and then stitched to obtain an image with a larger field of view. We introduce the notion of multispectral mosaicing in which much more information about the scene is extracted, by acquiring multiple images of the same scene by different sensors. The objective of this paper is to detail how mosaicing can be used for enhancement of the spectral information. First, we develop a geometric relationship between any two cameras.

Now a days various kind of sensors have been made and widely used in industries. However, due to limited resources of material, all those sensors can be made only sensitive to certain spectral band. For example, CCD cameras are designed for collecting visual signals, infrared sensors for measuring temperature (range $2 - 5\mu m$). Optical images from Landsat provide information on chemical composition, vegetation, and biological properties of the surface. In many cases, some ambiguities are caused when we use only one kind of sensor to perceive the real world.

In many applications, it is necessary to combine multiple images of the same scene acquired by different sensors, which often provide complementary information about the scene surveyed. For example, one could consider panoramic images of a house - both in the visual band, as well as in the IR band. The latter would be helpful, for example, to check if there has been any seepage in the walls.

Registered multisensor images can be directly fused at the pixel level and subsequent operations such as target detection and target recognition can be carried out using fused images instead of the individual sensor images. This not only saves the computations but also increases the target detection accuracy and target recognition rate. Schechner [8] describes wide field of view (FOV) multispectral imaging. A limitation of their approach is the use of special types of filters attached to the camera. Multispectral data is obtained in an extended FOV, using pushbroom [6]imaging spectrographs, which are generally rather complex and expensive. In [5], the authors register images (visible, IR) from Landsat and Spot satellites using a contour-based approach. A disadvantage of the above approach is the high computational complexity associated with their feature extraction, as well as registration processes.

In this paper, we develop a geometric relationship between panoramic mosaics in the visible and IR bands, corresponding to the same scene being imaged. Next, we describe a computationally efficient procedure to speed up panoramic mosaicing. In Section 4, we show results in support of the proposed strategies:inter-band superposition of panoramic mosaics.

2. Geometric Relationship

Two camera positions are related by a 3-D Euclidean transformation:

$$\mathbf{P}' = \mathcal{R}\mathbf{P} + \mathcal{T} \qquad (1)$$

Here, $\mathbf{P} = [X\ Y\ Z]^T$ and $\mathbf{P}' = [X'\ Y'\ Z']^T$ represent the (non-homogeneous) 3-D coordinates of a point viewed by the two camera stations, and let $\mathbf{p} = [x\ y\ 1]^T$ and $\mathbf{p}' = [x'\ y'\ 1]^T$ be the corresponding image points.

$$\mathcal{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix}$$

$$\begin{bmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

represent rotation and $\mathcal{T}=[T_x\ T_y\ T_z]^T$ translation between the world coordinate system and camera coordinate system. The 2-D image points and 3-D points in the camera coordinate system are related by

$$\lambda \mathbf{p} = \mathbf{AP} \qquad (2)$$

$$\lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \qquad (3)$$

where $\mathbf{A}$ represent the matrix of internal camera parameters. Here λ represents a projective constant, f_x and f_y represent the focal lengths in the x- and y- directions, s a skew factor and (u_0, v_0) represent the position of the principal point [4]. For the second image $\lambda'\mathbf{p}' = \mathbf{A}'\mathbf{P}'$ The above equation can be written as

$$\mathbf{p}' = \frac{\lambda}{\lambda'}(\mathbf{A}'\mathcal{R}\mathbf{A}^{-1})\mathbf{p} + \frac{1}{\lambda'}\mathbf{A}'\mathcal{T} \qquad (4)$$

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} X_1 & X_2 & X_3 \\ X_4 & X_5 & X_6 \\ X_7 & X_8 & X_9 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \qquad (5)$$

$$X_7 = \frac{1}{f_x}\left[\sin\alpha\sin\gamma - \cos\alpha\sin\beta\cos\gamma\right];$$

$$X_8 = \frac{1}{f_y}\left[D_3 - sX_7\right];$$

where $D_3 = (\sin\alpha\cos\gamma - \cos\alpha\sin\beta\sin\gamma)$.

$$X_9 = \left[\frac{-u_0}{f_x} + \frac{sv_0}{f_xf_y}\right]X_7 - \frac{v_0}{f_y}D_3;$$

$$X_1 = \frac{f_x'}{f_x}\cos\beta\cos\gamma - sD_1 + u_0'X_7;$$

where $D_1 = \frac{1}{f_x}\left(\sin\alpha\sin\beta\cos\gamma + \cos\alpha\sin\gamma\right)$.

$$X_2 = \frac{f_x'}{f_y}\left(\frac{s}{f_x}\cos\beta\cos\gamma + \cos\beta\sin\gamma\right)$$

$$\frac{s'}{f_y}(sD_1 + D_2) + u_0'X_8;$$

where $D_2 = (\cos\alpha\cos\gamma - \sin\alpha\sin\beta\sin\gamma)$.

$$X_3 = f_x'(\frac{-u_0}{f_x} + \frac{sv_0}{f_y})\cos\beta\cos\gamma$$

$$\frac{-v_0f_x'}{f_y}\cos\beta\sin\gamma + f_x'\sin\beta + u_0'X_9 +$$

$$s(\frac{-u_0}{f_x} + \frac{sv_0}{f_x})D_1 + \frac{-v_0}{f_y}D_2 + \sin\alpha\cos\beta;$$

$$X_4 = f_yD_1 + v_0'X_7;$$

$$X_5 = f_y'(\frac{-sD_1}{f_xf_y} + \frac{D_2}{f_y}) + v_0'X_8;$$

$$X_6 = f_y'D_1\left(\frac{-u_0}{f_x} + \frac{sv_0}{f_xf_y}\right) + \frac{-v_0D_2f_y'}{f_y} + \sin\alpha\sin\beta;$$
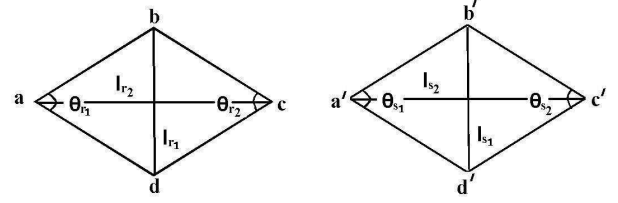

Figure 1: (a, b, c, d) basis quadruplet in reference image(left) and $(a', b',' c,' d')$ basis quadruplet in second image.

where primed quantities are internal camera parameters for the second image.

2.1. Special cases

We can reduce this general case to two cases of great practical significance. In most cases, one uses the same camera movement mechanism for both cameras. If the object of interest being mosaiced are very far away from the camera, the 3-D translation between the camera centers can be considered negligible with respect to the scene. Hence $b_i \approx 0$ for $1 \leq i \leq 3$. Hence, the relation between two corresponding points in two images reduces to a homography, Thus Equation 5 reduces to $\mathbf{p}' = \mu\mathcal{H}\mathbf{p}$. A further constraint can be put on this expression when two pairs of angles between the mosaicing cameras are zero namely α and β, and β and γ. In such a case, the last row of the homography $\mathbf{H} = [X_7\ X_8\ X_9]$ reduces to $[0\ 0\ 1]$, indicating an **affine transformation**. Such a situation arises for example, for panoramic mosaicing, when a camera rotates about the $Z-$ axis on a tripod. Thus two panoramas are related by a 2-D affine transformation.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} a_5 \\ a_6 \end{bmatrix} \qquad (6)$$

3. Efficient Panoramic Mosaicing

In the case of a collection of images of 3-D scene taken from the same point of view, the transformation between the images is linear transformation of 2-D projective space $\mathcal{P}^2$, called a collineation or a homography [9]. A commonly used camera model relating a world point $\mathbf{P_w}$ in $\mathcal{P}^3$, to its corresponding image point $\mathbf{p}$ in $\mathcal{P}^2$ is [4]:

$$\lambda\mathbf{p} = \mathbf{A}\left[\ \mathbf{R}\,|\,\mathbf{t}\ \right]\mathbf{P_w} \qquad (7)$$
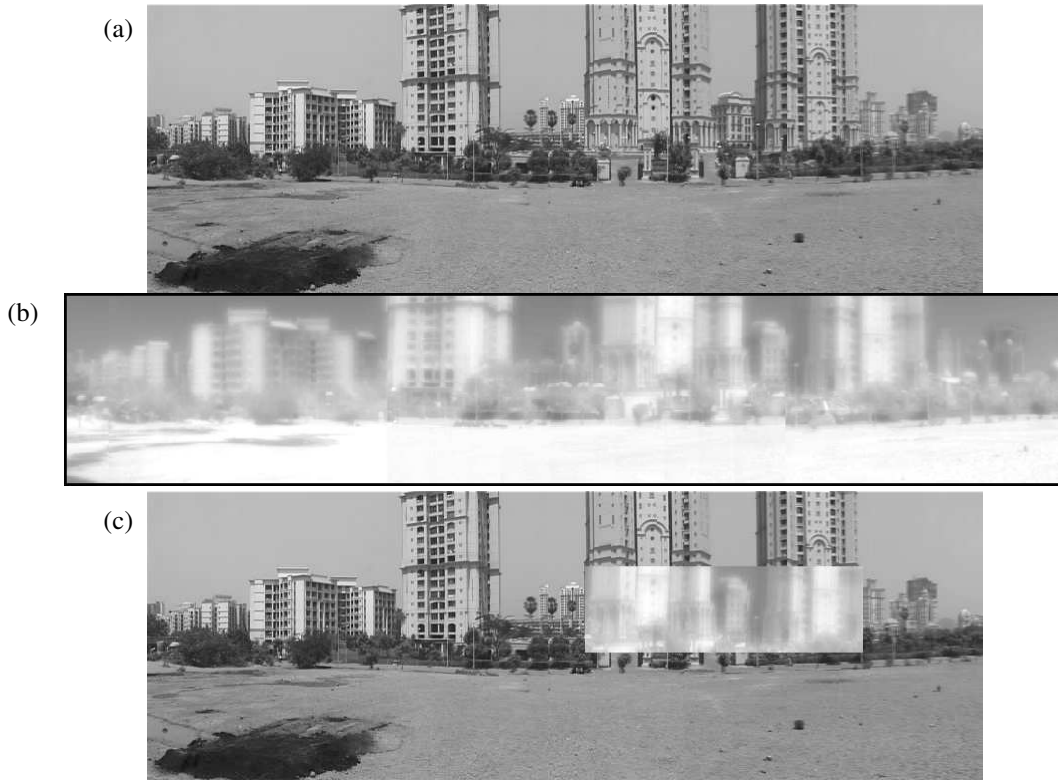
(a)

(b)

(c)

Figure 2: (a) A Panoramic visual image mosaic of the Hiranandani Complex, Mumbai, (b) the corresponding Panoramic IR image mosaic, and (c) Inter-band mosaic superposition of IR on visual band data

As in Section 2, we may write $\lambda_i \mathbf{p_i} = \mathbf{A_i P}$ and $\lambda_j \mathbf{p_j} = \mathbf{A_j P}$, for a 3-D point $\mathbf{P}$ viewed at two positions of the camera, i and j. The (non-homogeneous) 3-D coordinates of the world point, as viewed by the camera at the two positions, are related as

$$\mathbf{P_i} = \mathbf{RP_j} + \mathbf{t} \qquad (8)$$

For a panoramic imaging set up, the two camera positions have a negligible translation *i.e.*, $\mathbf{t} \approx \mathbf{0}$. Thus $\lambda_i \mathbf{A_i'}^{-1} \mathbf{p_i} = \lambda_j \mathbf{RA_j}^{-1} \mathbf{p_j}$. Hence, we have

$$\mu \mathbf{p_i} = \mathbf{Hp_j} \qquad (9)$$

where $\mathbf{H}$ is a 3×3 invertible, non-singular homography matrix. Homographies and points are defined up to a nonzero scalar. Every point correspondence gives two equations, thus to compute $\mathbf{H}$ (8 parameters), we need a four-point correspondence. We consider projective bases defined by pairs of four non-collinear projective points, using the canonical frame construction of [7]. We use a novel geometric hashing technique [2, 1, 3] for matching features in two images. The method reduces the exponential time complexity associated with the matching process, to a polynomial-time one, subject to the underlying transformation. Additionally, the papers [2, 1, 3] make an important observation: for a mosaicing application, the relative change of successive camera positions is often kept small to maximize the number of

corresponding points between images. For every quadruplet, we find angles formed by two linearly independent vectors and lengths between two end points as shown in in Figure 1. For every quadruplet in the second image, we find the difference between angle $\theta_{s_{1_j}}$ and angle $\theta_{r_{1_i}}$ and the difference between $\theta_{s_{2_j}}$ and the angle $\theta_{r_{2_i}}$ of all quadruplets in the reference image: $\delta_{\theta_{1_{(i,j)}}} = |\ \theta_{s_{1_j}} - \theta_{r_{1_i}}\ |$,
$\delta_{\theta_{2_{(i,j)}}} = |\ \theta_{s_{2_j}} - \theta_{r_{2_i}}\ |$ Similarly, we calculate the difference in lengths $\delta_{l_{1_{(i,j)}}}$ and $\delta_{l_{2_{(i,j)}}}$, where $i = 1, 2, 3...$ $\binom{M}{4}$; $j = 1, 2, 3... \binom{N}{4}$. Out of $\binom{M}{4} \times \binom{N}{4}$ combinations, the most likely correct pairs can be identified through two passes [2, 1, 3]. So, the pair with least values of $\delta_{\theta_1}, \delta_{\theta_2}, \delta_{l_1}, \delta_{l_2}$, considered as right candidate. Even though angles and lengths are *not* invariant parameters, they can be safely used as the relative change in these parameters is very small due to dense time-sampling of images. We use this information to estimate the transformation in the least-squares sense, to stitch together each pair of images.

4. Results and Discussion

Figure 2 shows the visual and IR panoramic mosaics and Inter-band mosaic of the Hiranandani Complex, Mumbai, describing an approximately 240° field of view. Figure 3 shows similar results for a school building. In both
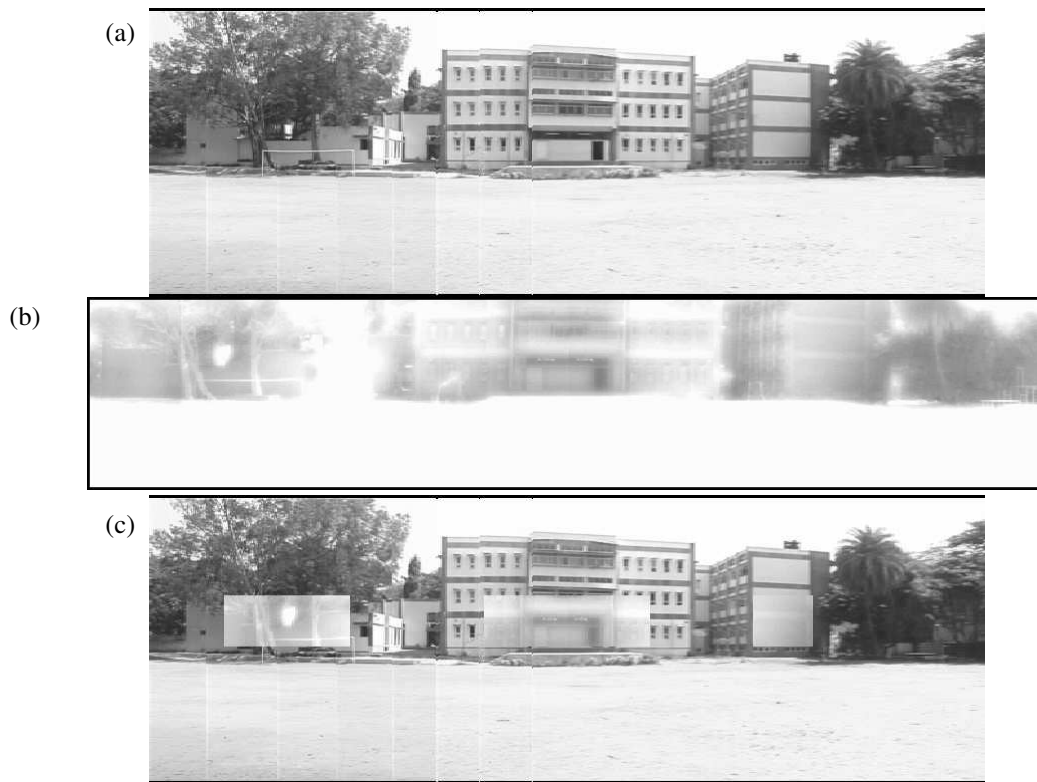
(a)

(b)

(c)

Figure 3: (a) A Panoramic visual mosaic of Central school, IITB, Mumbai, (b) The corresponding Panoramic IR image mosaic, and (c) Inter-band mosaic super position of IR on Visual band data

cases, we perform multispectral registration in an interactive, semi-automatic manner. Here, some of interest points are manually selected in both the IR and visual mosaics. By finding the correspondence between the two mosaics, we estimate the required affine transformation (Section 2.1) in a least-squares sense. Once the IR mosaic and visual mosaic are registered through the transformation, they are combined to form a multispectral mosaic. In our interactive system, any part of interest from IR mosaic can be obtained by just clicking points in the visual mosaic, and vice versa. Hence, we get multispectral information about the scene.

5. Conclusion

Here we present Multispectral Image mosaicing, which gives information about different modalities into scene, which is not possible with single camera. The registered multisensor images can be fused at the pixel level and subsequent operations can be carried out. This not only saves computations, it increases accuracy because the subsequent operations benefits from the spectral and geometric differences brought out by the fusion operation.

References

[1] U. Bhosle, S. Chaudhuri, and S. Dutta Roy. A Fast Method For Image Mosaicing Using Geometric Hashing. *IETE Journal of Research: Special Issue On Visual Media Processing*, pages 317 – 324, May - August 2002.

[2] U. Bhosle, S. Chaudhuri, and S. Dutta Roy. The Use of Geometric Hashing for Automatic Image Mosaicing. In *Proc. National Conference on Communication (NCC02)*, pages 533 – 537, January 2002.

[3] U. Bhosle, S. Chaudhuri, and S. Dutta Roy. Background Mosaicing of Scene With Moving Objects. In *Proc. National Conference of Communication (NCC03)*, pages 84 – 89, January 2003.

[4] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[5] L. Hui. Automatic Visual/IR Image Registration. *Optical Engineering*, 35(2):395 – 400, 1996.

[6] J.Wellman. Multispectral Mapper:Imaging Spectroscopy as Applied to the Mapping of Earth Resources. In *Proc. SPIE Imaging Spectroscopy*, volume 268, pages 64 – 73, 1981.

[7] C. A. Rothwell. *Recognition using Projective Invariance*. PhD thesis, University of Oxford, 1993.

[8] Y. Schechner and S. Nayar. Generalized Mosaicing:Wide Field of View Multispectral Imaging. *IEEE Trans. on Pattern Anal. and Machine Intell.*, 24(10):1334 – 1349, October 2002.

[9] I. Zoghlami and R. Deriche. Using Geometric Corners to Build a 2D Mosaic from a Set of Images. In *Proc. IEEE International Conference on Image Processing (ICIP)*, pages 420 – 425, March 1997.