Multispectral Panoramic Mosaicing

Udhav Bhosle, Sumantra Dutta Roy and Subhasis Chaudhuri
Department of Electrical Engineering
Indian Institute of Technology Bombay, Mumbai -400 076 (India)
{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 suitably transformed frames in a sequence of overlapping images. This paper deals with multispectral mosaicing for enhancement of spectral information – information from the thermal infra-red (IR, hereafter) band and visible band images is directly fused at the pixel level. All subsequent operations can be carried out using the fused mosaic instead of the individually sensed images. We develop a geometric relationship between a visible band panoramic mosaic and an IR one. Our system uses a fast algorithm for automatic panoramic mosaicing. We show results of inter-band mosaic superposition in support of the proposed strategies.

Key words: Mosaicing, multispectral mosaicing, Geometric Hashing, least-squares estimation.

1 Introduction

The general problem of mosaicing is to create a single seamless image by aligning a series of spatially overlapped images. The result is an image with a field of view (FOV) greater than that of a single image. Traditional image mosaicing mainly addresses the extension of FOV, while other imaging dimensions are not improved in this process. The objective of this paper is to detail how mosaicing can be used for enhancement of the spectral information.

Now-a-days various kinds of sensors have been made and are widely used in the industry. However, due to limited resources of material, sensors can be made only sensitive to certain signals. For example, CCD cameras are designed for collecting visual signals, thermal infrared sensors for measuring temperature, electro-magnetic induction sensors for metal detection. Optical images from

Landsat provide information on chemical composition, vegetation, and biological properties of the surface. In most cases the information provided by each sensor may be incomplete, inconsistent or imprecise. In many cases, some ambiguities will be caused when we use only one kind of sensor to perceive the real world. Additional sources may provide complementary data in addition to the redundant information content. Merging of redundant data can reduce imprecision, and fusion of complementary data can create a more consistent interpretation. Therefore, it is useful to combine and analyze the multi-source data to take advantage of their characteristics and enhance the information extraction process. Besides overcoming the ambiguity problem, fusion can also bring some benefits to our perception of 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 being surveyed. For example, one can consider panoramic images of a house - both in the visible band, as well as in the thermal IR band. The latter would be important for example, to check for seepage in the walls. In a military application for example, soldiers and tanks could stand camouflaged with the background. Having registered images from a visible band camera and a thermal IR one, it helps one to identify the camouflaged entities in relation to their surroundings. The images from different sensors are registered, and these images can be directly fused at the pixel level and subsequent operations such as target detection and recognition can be carried out using the fused images instead of the individual sensor image. This saves on computation, and increases target detection accuracy and recognition rate because subsequent operations benefit from spectral and geometric differences brought out by fusion operations.

Automatic registration of visible band and thermal IR data is very difficult because of the inconsistent features present in the data, as features present in thermal IR images are not often same as those in the visible band images. If the images are similar, as in those formed from similar sensors, one could rely on correspondence on gray levels or texture for registration. However, in images taken with sensors operating on different spectral bands (for instance visible band and thermal IR images) texture and gray-levels do not often match. In some special cases, as in registration of medical images from different sensors, contextual considerations give additional information about the images to register, but these cannot be generalized for the registration of real world scenes.

Schechner and Nayar [1] describe wide 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 push-broom imaging spectrographs [2] which are generally rather complex and expensive. Rignot *et al.* [3] describe a conceptual approach that integrates a variety of

registration techniques and selects the candidate algorithm based on certain performance criteria. The performance requirements for an operational algorithm are formulated given the spatially, temporally and spectrally varying factors that influence the image characteristics and the requirement of various applications. The authors use a feature-based method for registration. The authors use the binary correlation and chamfer matching technique for matching the features. This technique is not applicable for registration of all types of sensors. In [4] and [5], the authors register images 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 the registration processes.

The simplest forms of mosaics are created from a set of images whose mutual displacement are pure image plane translation. This is approximately the case with satellite images. Such translation can either be computed by manually pointing to corresponding points or by an image correlation methods [6]. Other simple mosaics are created by rotating the camera about its optical center, using a special device and creating a panoramic image, which represents the projection of the scene onto a cylinder [7,8]. Since it is not simple to ensure a pure rotation around the optical center, such mosaics are used only in limited cases. For large rotation around the optical axis, a few methods have been developed in the literature. An efficient scheme is that of Dani and Chaudhuri [6], which utilizes the angular histogram for rotation estimation. Their method works up to 15° rotation. Some efficient methods have also been developed to build a mosaic, when the homography is mainly translation. For example, if the overlap between the images is very large (*i.e.*, the motion is very small), it has been shown that the Levenberg-Marquardt method yields good results [7], but it is very sensitive to local minima and is computationally expensive. When the overlap is small, one can use a hierarchical matching to avoid local minima. For large camera motion, the authors in [9] propose a phase correlation method. There has been a substantial research at Sarnoff [10], that aims at obtaining a robust panoramic mosaic. This works well for video data when the images in successive frames have similar photometric properties. Since we concentrate on multispectral data, the photometric information cannot always be used for matching purposes. Hence we restrict ourselves to using only feature points (geometric properties) for mosaicing purposes.

In this paper, we first develop a geometric relationship between panoramic mosaics in the visible and thermal IR bands, corresponding to the same scene being imaged (Section 2). To construct a panoramic mosaic in an efficient manner, Section 3 describes a Geometric Hashing-based scheme. Section 4 describes our multispectral mosaicing system. In Section 5, we show results in support of the proposed strategy: inter-band superposition of panoramic mosaics.

2 Geometric Relationship

A commonly used basic perspective projection model is that of a pin-hole camera [11], which describes the relationship between a 3-D world point and its corresponding 2-D image point:

$$\lambda \mathbf{p} = \mathbf{A}\,[\mathbf{R}|\mathbf{t}]\,\mathbf{P}_W \tag{1}$$

Here $\mathbf{p}$ is a 3×1 vector of image coordinates in $\mathcal{P}^2$ ($\mathcal{P}^2$ represents 2-D projective space) corresponding to a 3-D world point, where 3-D coordinates with respect to (an arbitrary) world coordinate system are given by the 4×1 vector $\mathbf{P}_W = [X_W\ Y_W\ Z_W\ H]^T$. $H = 0$ represents an ideal 3-D point (point on the plane at infinity). For a real 3-D point, we may consider $H = 1$, without loss of generality. Let $\widetilde{\mathbf{P}}_W = [X_W\ Y_W\ Z_W]^T$ represent the corresponding non-homogeneous coordinates. $\mathbf{R}$ and $\mathbf{t}$ represent the external camera parameters relating the camera coordinate system to the world coordinate system, and $\mathbf{A}$ is the matrix of camera internal parameters given by

$$\mathbf{A} = \begin{bmatrix} f_x & s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Here f_x and f_y are the camera focal lengths in the $x-$ and $y-$ directions, respectively; and u_0 and v_0 represent the position of the principal point. s is the skew factor- this can be assumed to be negligible [12]. λ is a projective constant.

Equation 1 can be written as

$$\widetilde{\mathbf{P}}_W = \mathbf{R}^{-1}\mathbf{A}^{-1}\,[\lambda \mathbf{p} - H\mathbf{A}\mathbf{t}]\,. \tag{2}$$

For another camera (with the corresponding variables denoted as primed quantities);

$$\widetilde{\mathbf{P}}_W = \mathbf{R}'^{-1}\mathbf{A}'^{-1}\,[\lambda' \mathbf{p}' - H\mathbf{A}'\mathbf{t}'] \tag{3}$$

Equating the two expressions in Equations 2 and 3, we get

$$\lambda' \mathbf{p}' = \lambda \mathbf{A}' \mathcal{R} \mathbf{A}^{-1} \mathbf{p} + H\mathbf{A}'\mathcal{T} \tag{4}$$

where $\mathcal{R} = \mathbf{R}'\mathbf{R}^{-1}$ and $\mathcal{T} = \mathbf{t}' - (\mathbf{R}'\mathbf{R}^{-1})\mathbf{t}$. $\mathcal{R}$ and $\mathcal{T}$ represent the 3-D Euclidean relationship between the two camera coordinate systems. Here $\mathbf{A}'$ represents the internal camera parameter matrix for the second camera.

Equation 4 can be written in a matrix form as

$$\lambda'\mathbf{p}' = \mathcal{H}\mathbf{p} + H\mathbf{A}'\mathcal{T} \tag{5}$$

The exact form of the $\mathcal{H}$ matrix in terms of the camera parameters can be found in [13]. From Equation 5, points $\mathbf{p}'$ and $\mathbf{p}$ are related by a homography when

- The world point is an ideal point *i.e.*, $H = 0$, and
- When two camera positions have no translation between them *i.e.*, have the same camera optical centre.

Under these assumptions, Equation 5 reduces to

$$\lambda'\mathbf{p}' = \mathcal{H}\mathbf{p} \tag{6}$$

Hence, points $\mathbf{p}$ and $\mathbf{p}'$ are related by a homography. Equation 6 has 8 independent parameters.

2.1 Special Cases

We can reduce this general case (Equation 5) to three cases of great practical significance.

- In most cases, one uses the same camera movement mechanism for both cameras (*e.g.*, one shown in Figure 1). If the objects 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. Here again, we get the relation between $\mathbf{p}'$ and $\mathbf{p}$ to be a homography. This can be seen quite easily *i.e.*, without loss of generality, one may consider the world coordinate system origin to be situated close to the objects being imaged (basically, very far from the cameras). The (non-homogeneous) coordinate of the two camera centres in their own coordinate systems are $[0\ 0\ 1]^T$. Using $\mathbf{P} = \mathbf{R}\widetilde{\mathbf{P}}_W + \mathbf{t}$, one obtains the coordinate of the first camera centre in the world coordinates system as $-\mathbf{R}^{-1}\mathbf{t}$. The corresponding quantity in the second camera's coordinate system is $-\mathbf{R}'^{-1}\mathbf{t}'$. Since the distance between the two cameras is negligible with respect to the distance from the world coordinate system origin, their world coordinates are nearly the same *i.e.*,

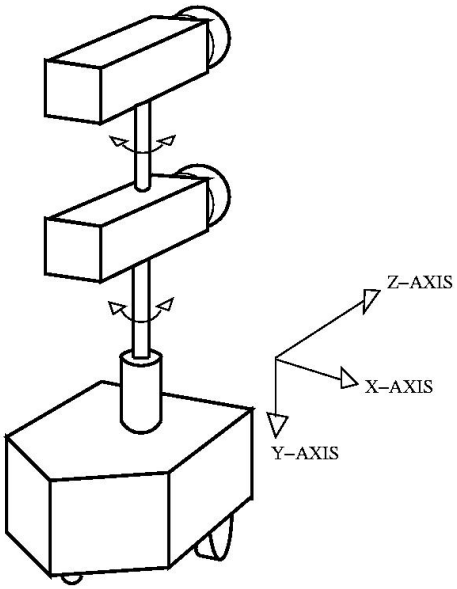

Fig. 1. Experimental set up for capturing the multi-sensor panoramic data.

$-\mathbf{R}^{-1}\mathbf{t} \approx -\mathbf{R}'^{-1}\mathbf{t}'$, or $\mathbf{t}' - \mathcal{R}\mathbf{t} \approx \mathbf{0}$. Thus $\mathcal{T} \approx \mathbf{0}$. Hence, the relation between two corresponding points in two images reduces to a homography.

- An alternate constraint can be put on Equation 5 when the Euler angles with respect to $\mathbf{X}-$ and $\mathbf{Y}-$ coordinates are negligible. In such a case, the last row of the homography $\mathcal{H} = [h_7\ h_8\ h_9]$ reduces to [0 0 1], indicating an **affine transformation**. Such a situation arises for example, for panoramic mosaicing, when a camera rotates about the $\mathbf{Z}-$ axis on a tripod. Thus the two panoramas are related by a 2-D affine transformation.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} a_5 \\ a_6 \end{bmatrix}. \tag{7}$$

- Consider $\mathcal{T} \neq 0$ and assume that the 3-D point is a real point *i.e.*, $H = 1$. Equation 4 now takes the form

$$x' = \frac{h_1 x + h_2 y + h_3 + t_1}{h_7 x + h_8 y + 1 + t_3}; \quad y' = \frac{h_4 x + h_5 y + h_6 + t_2}{h_7 x + h_8 y + 1 + t_3} \tag{8}$$

where the t_i terms arise from the $H\mathbf{A}'\mathcal{T}$ term in Equation 4, and h_9 can be taken to be 1, without loss of generality. The above transformation will be affine only if $h_7 = h_8 = 0$.

3 Efficient Panoramic Mosaicing

In the case of a collection of images of a 3-D scene taken from the same point of view, the transformation between the images is a linear transformation of 2-D projective space $\mathcal{P}^2$, called a collineation or a homography [14].

Equation 1 gives the basic projection model of a camera. For two viewing positions i and j, we can write the relationship between the image points $\mathbf{p}_i$ and $\mathbf{p}_j$ in terms of the 3-D coordinates with respect to the camera coordinate systems at position i and j as $\lambda_i \mathbf{p}_i = \mathbf{A}_i \mathbf{P}_i$ and $\lambda_j \mathbf{p}_j = \mathbf{A}_j \mathbf{P}_j$, respectively. Further, just as in Section 2, the two positions i and j are related by a 3-D Euclidean transformation $\mathbf{P}_j = \hat{\mathbf{R}}\mathbf{P}_i + \hat{\mathbf{t}}$.

For a panoramic imaging set up, the two camera positions have a negligible translation between them *i.e.*, $\hat{\mathbf{t}} \approx \mathbf{0}$. Thus $\lambda_j \mathbf{A_j^{-1}} \mathbf{p_j} = \lambda_i \mathbf{R} \mathbf{A_i^{-1}} \mathbf{p_i}$. Hence, we have

$$\mu \mathbf{p_j} = \hat{\mathbf{H}} \mathbf{p_i} \tag{9}$$

$$\mu \begin{bmatrix} x_j \\ y_j \\ 1 \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \tag{10}$$

where $\hat{\mathbf{H}}$ is a 3×3 invertible, non-singular homography matrix. Homographies and points are defined up to a nonzero scalar. For the principal point of image 1 we have $[x_{i_1}, y_{i_1}] = [0, 0]$. Its corresponding location in the coordinates of image 2 is $[\frac{h_3}{h_9}, \frac{h_6}{h_9}]$. As long as the camera is well above the ground, the principal point of image 1 must be well defined point (finite) in the coordinates of image 2. Hence $h_9 \neq 0$. Without loss of generality, we may take $h_9 = 1$. The above equation can be written as

$$x_j = \frac{h_1 x_i + h_2 y_i + h_3}{h_7 x_i + h_8 y_i + 1}; \quad y_j = \frac{h_4 x_i + h_5 y_i + h_6}{h_7 x_i + h_8 y_i + 1}. \tag{11}$$

Every point correspondence gives two equations, thus to compute $\hat{\mathbf{H}}$ (8 parameters), we need four-point correspondences. For a pair of corresponding points, this can be written as

$$\begin{aligned} h_1 x_i + h_2 y_i + h_3 - h_7 x_i x_j - h_8 y_i y_j &= x_j \\ h_4 x_i + h_5 y_i + h_6 - h_7 x_i y_j - h_8 y_i y_j &= y_j. \end{aligned}$$

Let us assume that the two images to be registered have M and N feature points, respectively. We use a novel Geometric Hashing-based technique [15,16] for matching the features in two images. The method reduces the exponential time complexity associated with the matching process to a polynomial-time one, by not performing an exhaustive search. A 2-D transformation requires K basis points ($K = 3$ for Euclidean and affine, 4 for projective). We can select ordered pairs of K basis points from the first image in $\binom{M}{K} \times K!$ ways (this is $\mathcal{O}(M^K)$). A *hash table* stores these coordinates, indexed by the basis points. We repeat the process for the second image. Matching rows of coordinates between hash tables of the two images has a *quadratic* time complexity. We can reduce this to a *linear* form if we sort each row in the hash tables. Hence the problem of matching image features reduces to $\mathcal{O}(M^{K+1}N^{K+1}) \times$ *the row matching time*. This has polynomial time complexity, an improvement over the exponential time complexity required, as for a naive feature match. This is in a restricted case - when we assume a specific form of linear transformation between the two points sets.

We consider projective bases defined by pairs of four non-collinear projective points, using the canonical frame construction of [17]. This method considers mapping from the four non-collinear points to the corners of a unit square. Thus we have $\binom{M}{4} \times M!$ possible choices for the basis vectors. We can often make a further assumption to reduce the matching time complexity. *The relative change of successive camera positions is often kept small, so as to maximize the region of overlap between images* - to maximize the number of corresponding points. Here, we have taken the angle θ_1 and θ_2 (see Figure 2) formed by two linearly independent vectors based on these basis quadruplets, and length l_1 and l_2 between the two end points as parameters in the hash table. Thus, even though lengths and angles are *not* projectively invariant, the above constraints can be safely used while matching, as the relative change in these parameters is very small due to the dense time sampling of images. So, the mosaicing algorithm can be succinctly given as

Algorithm:

- Represent the reference image by a set of features.
- For every quadruplet (of which three must be non-collinear), find the angles (θ_1, θ_2) formed by two linearly independent vectors and lengths (l_1, l_2) between the two end points as shown in Figure 2.
- For the second frame of the scene, for every quadruplet, find the corresponding (θ, l) values.
- For every quadruplet in the second image find the difference between the angles $\theta_{s_{1_j}}$ and $\theta_{r_{1_i}}$, and the difference between $\theta_{s_{2_j}}$ and $\theta_{r_{2_i}}$ for all quadruplets in the reference image.

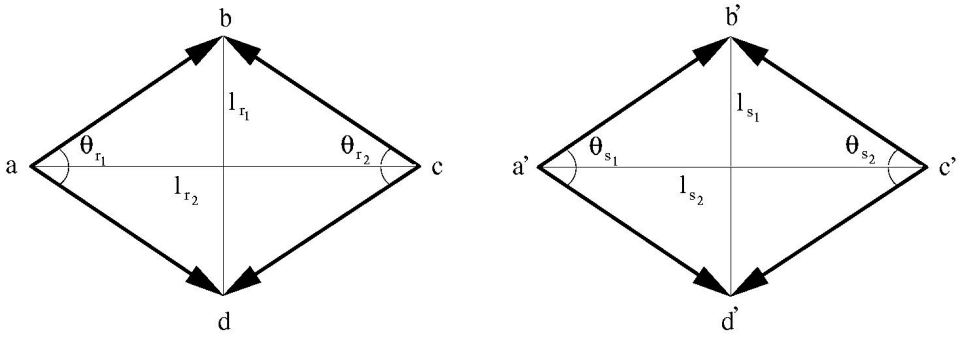

Fig. 2. (a, b, c, d) basis quadruplet in reference image (left) and $(a', b,' c,' d')$ basis quadruplet in second image (right).

The difference in angles can be calculated as

$$\delta_{\theta 1_{(i,j)}} = \mid \theta_{s1_j} - \theta_{r1_i} \mid; \quad \delta_{\theta 2_{(i,j)}} = \mid \theta_{s2_j} - \theta_{r2_i} \mid$$

Here subscripts r and s refer to the reference image and the other image to be matched to it, respectively.

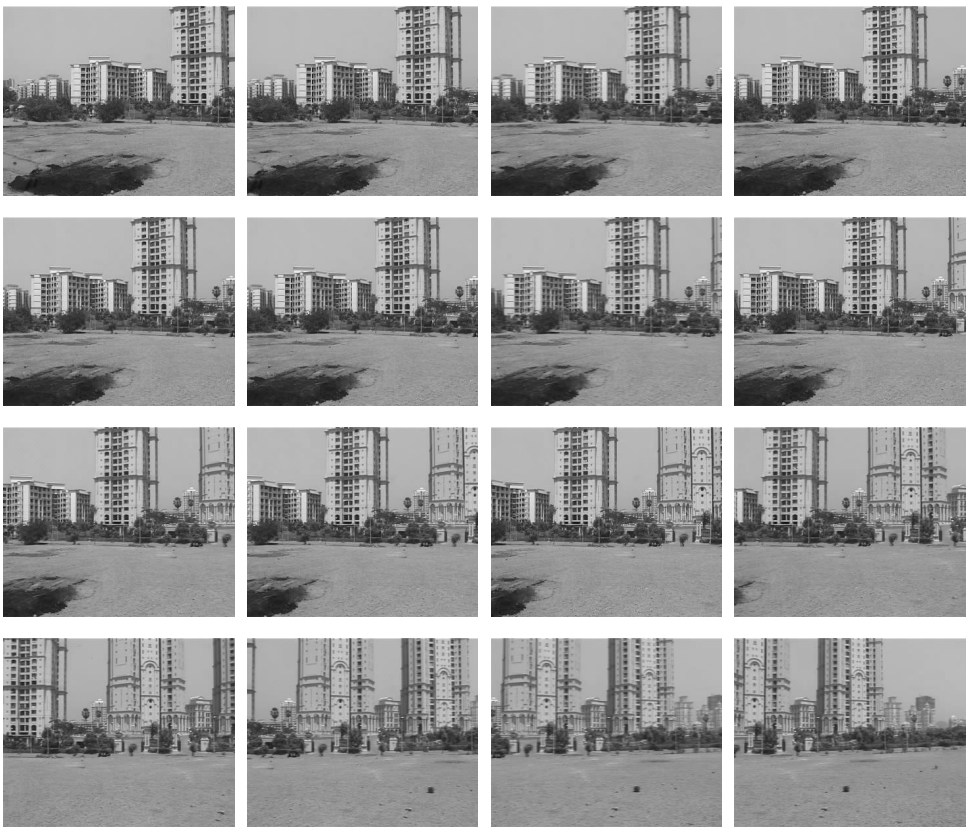

Fig. 3. Some images of the Hiranandani Complex, Mumbai taken with a visible band camera.

Similarly, we calculate the difference in lengths

$$\delta_{l_{1_{(i,j)}}} = \mid l_{s1_j} - l_{r1_i} \mid; \quad \delta_{l_{2_{(i,j)}}} = \mid l_{s2_j} - l_{r2_i} \mid$$

where $i = 1, 2, 3 \quad ... \binom{M}{4}$; $j = 1, 2, 3 \quad ... \binom{N}{4}$. Here, the M and N represent the number of feature points in the reference image and other image, respectively. Out of $\binom{M}{4} \times \binom{N}{4}$ combinations, only a few most likely correct pairs can be identified through two passes. At first quadruplets are compared based on angles. We can discard the quadruplets which give an angle difference more than a specific threshold. In our experiments, we typically observe that about 90% of the quadruplets get pruned out in this phase alone. The pairs of quadruplets with small difference in θ_1 and θ_2 (typically, 10% of the original set of quadruplets) will be considered for comparison based on lengths. By sorting based on δ_{l_1} and δ_{l_2}, we choose pairs with minimum value of δ_{l_1} and δ_{l_2}. So, the pair with least values of $\delta\theta_1, \delta\theta_2, \delta_{l_1}, \delta_{l_2}$ considered as the right candidate. So a quadruplet in the reference image matches with a quadruplet in the second image. This means that four points in the first image have correspondences with four points in the second image.

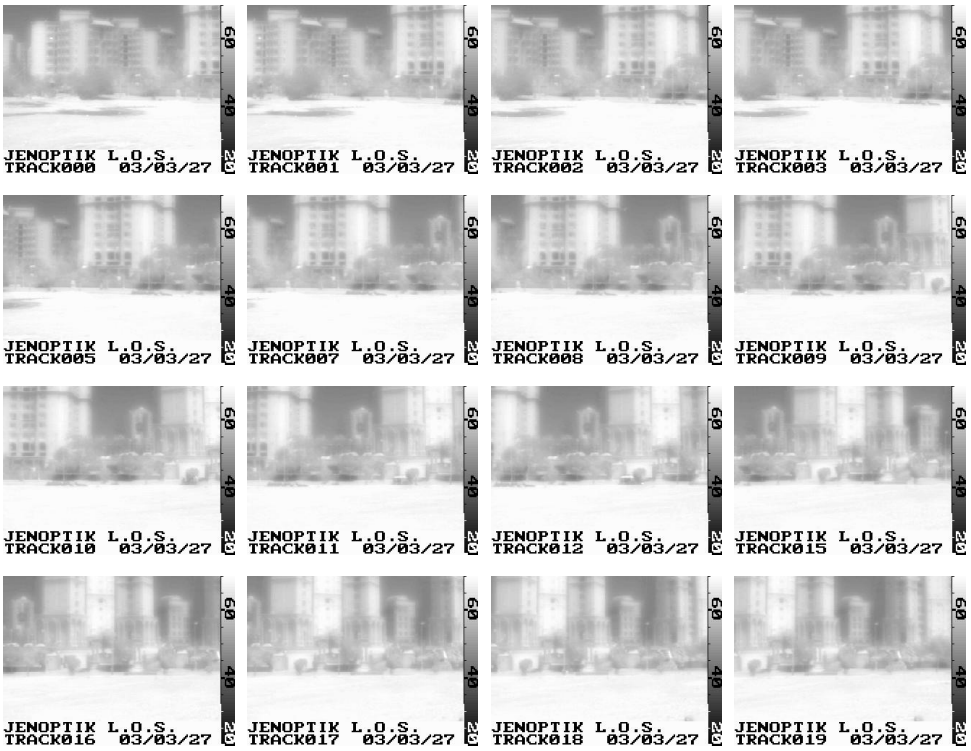

Fig. 4. Some images in the thermal IR band of the Hiranandani Complex, Mumbai. The scale to the right shows the temperature range.

In the above algorithm, for the reference image $\theta_1, \theta_2, l_1, l_2$ values are stored. In the matching part, while choosing any four points from the second image, depending on the values of $\theta_1, \theta_2, l_1, l_2$, there may not be any match. So in order

to have robustness in matching, every non collinear four points are stored in the hash table. Also in the second image, we look for only one quadruplet arbitrarily, and based on $\theta_1, \theta_2, l_1, l_2$, we find for a match in the table. Though it may match, it might be a wrong candidate and there may be some other quadruplet in the second image which can match with the quadruplet in first image. So, in order to avoid this ambiguity, a comparison is done for all the possible non-collinear quadruplets in the second image. Hence this method reduces the exponential time complexity associated with the matching process, to a polynomial-time one, subject to the underlying transformation between the point sets [15,16].

By knowing the correspondence, we can find the initial estimate of homography $\hat{\mathbf{H}}_{initial}$ between the images. Then $\hat{\mathbf{H}}_{final}$ is obtained using least squares estimation (LSE). First $\hat{\mathbf{H}}_{initial}$ is obtained from the matched quadruplets. We select a number of feature points in the reference image, centered around the basis quadruplet. These points are transformed according to $\hat{\mathbf{H}}_{initial}$ in the second image and we find the match for the same. In the second image, around the transformed point, we draw a circle with a radius of two pixels. If the transformed point is within the circle, then we say that the point in the reference image has a correspondence in the second image. Otherwise, we discard that point. In this way we continue for a number of points and set the correspondence for more points in the reference image and the second image. For more than four corresponding vertex pairs, say $((x_{i_1}, y_{i_1}), (x_{i_2}, y_{i_2}) \ldots (x_{i_h}, y_{i_h}))$ and $((x_{j_1}, y_{j_1}), (x_{j_2}, y_{j_2}) \ldots (x_{j_h}, y_{j_h}))$ where $h \geq 4$, the required transformation can be obtained as a solution which minimizes the least squares error

$$\epsilon = \sum_{k=1}^{h} \parallel \mathbf{p}_{j_k} - (\hat{\mathbf{H}}\mathbf{p}_{i_k}) \parallel^2 \tag{12}$$

with respect to the motion parameters. In the above expression, we consider non-homogeneous versions of 2-D image coordinates $\mathbf{p}_{j_k}$ and $\hat{\mathbf{H}}\mathbf{p}_{i_k}$. By estimating the transformation, the second image is transformed with respect to the first image and these images are combined to form a mosaic. Here the reference image is selected and all other images are aligned with respect to the first image, and they are combined to form a mosaic. In this case, overlapping area between a pair of images is taken only from one of the images, so there is no effect of blurring in the mosaic, but a visible seam may be present in the mosaic. One may need to use a suitable post processing technique for removal of seam in the visible band. The mosaics in the IR band does not usually suffer from a seam as the measurements indicate the temperature map and there is no automatic gain control here. In this work, we assume no moving objects in the scene. We examine the case of mosaics in scenes with moving objects in our earlier paper [16].

4 Multispectral Mosaic Matching

An attractive aspect of multi-sensor data acquisition is the fact that the sensors collect very different types of information from the same scene. Our system first constructs panoramic mosaics for both the visible band and the thermal IR band using the Geometric Hashing-based technique discussed in previous section. At this stage, we have the individual panoramas in different spectral bands (Section 3), and the geometric relationship between the two (Section 2). To obtain an inter-band mosaic, a user can select the required region from any mosaic. The system uses the above geometric relationship between the panoramas, to fill in the required region from the other panorama. Thus, registered multi-sensor images can be directly fused at the pixel level and subsequent operations can be carried out using the fused image instead of the individual sensor images. This not only saves computation time, but also increases the accuracy because the subsequent operations benefit from the spectral and geometric differences brought out by fusion operations.

The following points may be noted while constructing a multispectral mosaic. The accuracy of feature detection in all the bands need not be identical. For the visual band data, one can obtain a very good image contrast even with an inexpensive, commercially available camera. Thus, the localization of feature points is quite good and hence, the resulting panorama does not suffer from geometric distortion. On the other hand, thermal cameras do not offer a high-contrast picture. This is due to the fact that all objects try to attain a natural thermal equilibrium, and this is a continuous process. The feature detection and localization are less accurate in thermal IR bands and hence the mosaic may have a higher geometric distortion. The saving grace is that the geometric distortion in thermal band mosaics may not be always visually perceived since the images themselves have a poor contrast. Nevertheless, such a distortion would limit its applications in scene interpretation. One may also notice that we do not make any attempt to register the multiband data at the individual frame level. This is due to the fact that the photometric properties are very different in different bands, and the features may not always be detected over several bands with a sufficient accuracy. Hence we try to register the mosaics over different bands. This being a global process, we are more likely to pick up a few better feature points over the entire panorama, thus providing a more reliable overall registration.

5 Results and Discussion

We took an arbitrary set of images of the Hiranandani Complex, Mumbai using a panoramic set up. Figure 1 shows the experimental set up for capturing

multispectral images. To capture the images, a SONY HANDYCAM (DCR - TRV900E PAL) operating in the visible band was mounted on a level tripod and we operate it in the manually preset mode where all the camera parameters such as aperture, shutter speed, focal length were kept constant. Since the images are taken by revolving the camcorder in the scene, the rotation between the images is unknown and is not assumed to be constant. Figure 3 shows some images of the Hiranandani Complex, Mumbai, in the visible band. Next, images were also captured using a thermal IR camera (Jenoptic Varioscan thermal IR camera, range $2 - 5\mu m$) with the same set up. The spatial overlap for the successively captured images varied in the range of $50 - 70\%$. Figure 4 shows some of the images of the same scene in the IR band. First we construct the panoramic mosaics for both visible band and thermal IR images separately, using the Geometric Hashing-based technique, as discussed above. We select one image as the reference image and all other images are transformed according to the estimated transformation - they are stitched together to form a mosaic. Then we construct the panoramic mosaic in the thermal IR band. Figure 5 shows the visible band mosaic, the thermal IR panoramic mosaic, and super-posed mosaic of the Hiranandani Complex, describing a Field of view of approximately 240°. The accumulated error for the 240° field of view was found to be about 2 pixels for both the bands. One may notice from the panoramas that the tall buildings in the front are quite warm (about 60° - 65° C) as they were facing the sun. For the same reason, the asphalted road in the front is quite warm. The vegetation and the small buildings in the shade are relatively cooler (about 30°- 40° C). The above phenomenon can easily be seen from the super-posed mosaic in the Figure 5(c).

We now show the results of experiments on a different data set. Figures 6 and 7 show some images of the Central School, IIT Bombay, in the visible band and the thermal IR band, respectively. Figure 8 shows similar results of mosaicing for the school building. One can draw similar conclusion from the super-posed mosaic. For example, the metallic object above the goal post and toward the left middle of the mosaic is quite warm compared to the vegetation seen in the picture. In both cases, we perform multispectral registration in an interactive, semi-automatic manner. Here, some points of interest are selected in both the thermal IR and visible band mosaics. By finding the correspondence between the two mosaics, we estimate the required transformation (Section 2.1) in a least-squares sense. The computed matrix showing the transformation between the visible band and thermal IR band mosaic for the case of the Hiranandani Complex is

$$\mathcal{H} = \begin{bmatrix} 1.252 & -0.026 & 19.115 \\ 0.002 & 1.162 & -74.881 \\ 0.000 & 0.000 & 1.000 \end{bmatrix}.$$

(a)

(b)

(c)

Fig. 5. Multispectral panoramic mosaic of the Hiranandani Complex, Mumbai, (a) visible band panorama, (b) thermal IR band panorama, and (c) inter-band superposition of thermal IR data on the visible band panorama, at a given location.

Similarly, the following matrix was computed for the transformation between the visible band and thermal IR panoramic mosaic for the Central School.

$$\mathcal{H} = \begin{bmatrix} 1.332 & -0.023 & -25.849 \\ 0.003 & 1.158 & -81.388 \\ 0.000 & 0.000 & 1.000 \end{bmatrix}.$$

Once the thermal IR mosaic and visible band mosaic are registered through the transformation, they are combined to form a multispectral mosaic. In our

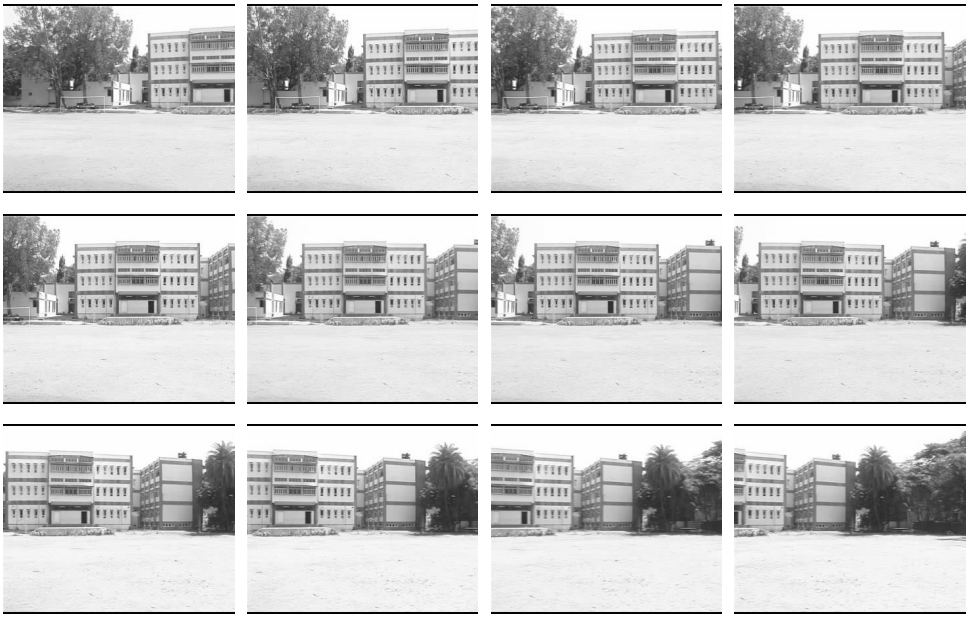

Fig. 6. Some images of a video sequence of the Central School, IIT Bombay in the visible band.

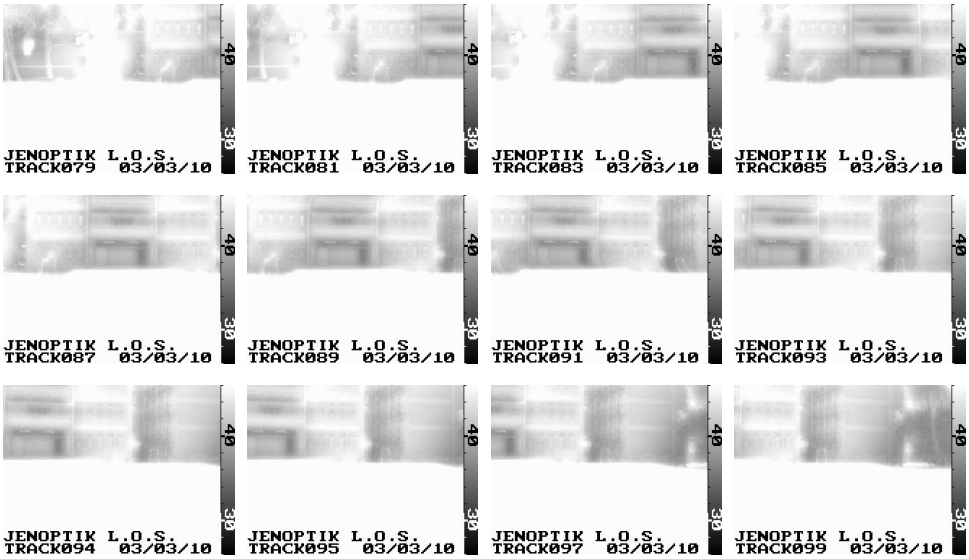

Fig. 7. Some images of the above Central School in the thermal IR band.

interactive system, any part of interest from the thermal IR mosaic can be obtained by just clicking on points in the visible band mosaic, and vice versa. Hence, we get the multispectral information about the scene.

(a)

(b)

(c)

Fig. 8. Multispectral panoramic mosaic of the Central School, serves as another example. (a) visible band panorama, (b) thermal IR band panorama, and (c) inter-band super-position of thermal IR data on the visible band panorama, at a given location.

6 Conclusion

We develop a method for multispectral image mosaicing, which gives information about different aspects of the scene, which is not possible with a single camera. The registered multi-sensor images can be fused at the pixel level and subsequent operations can be carried out on the fused mosaic. This not only saves computations, it increases accuracy because the subsequent operations benefit from the spectral and geometric differences brought out by the fusion operation. We achieve the goal by using simple sensors instead of using

a complicated and costly sensor. Our future research will include processing of the multispectral mosaics for feature classification and scene understanding purposes.

Acknowledgment

The authors are thankful to the reviewers for their constructive comments. Partial funding under the *Swarnajayanti* Fellowship scheme is acknowledged.

References

[1] Y. Schechner, S. Nayar, Generalized Mosaicing: Wide Field of View Multispectral Imaging, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (10) (2002) 1334 – 1349.

[2] J. Wellman, Multispectral Mapper: Imaging Spectroscopy as Applied to the Mapping of Earth Resources, in: Proc. SPIE Imaging Spectroscopy, Vol. 268, 1981, pp. 64 – 73.

[3] L. Rignot, R. Kowk, J. Curlander, S. Pang, Automated Multisensor Registration: Requirements and Techniques, Photogrammetric Engineering and Remote Sensing 57 (8) (1991) 1029 – 1038.

[4] L. Hui, B. Manjunath, A Contour Based Approch to Multisensor Image Registration, IEEE Transactions on Image Processing 4 (3) (1995) 320 – 334.

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

[6] P. Dani, S. Chaudhuri, Automated Assembly of Images: Image Montage Preparation, Pattern Recognition 28 (3) (1995) 431 – 445.

[7] R. Szeliski, H. Yeung, Creating Full View Panoramic Image Mosaic and Environment Maps, in: Computer Graphics Proceedings, Annual Conference Series, 1991, pp. 251 – 258.

[8] S. Peleg, B. Rousso, Universal Mosaicing Using Pipe Projection, in: Proc. IEEE International Conference on Image Processing (ICIP), 1998, pp. 123 – 132.

[9] L. Brown, A Survey of Image Registration Techniques, ACM Computing Surveys 4 (1992) 325 – 376.

[10] H. Sawhney, S. Hsu, R. Kumar, Robust Video Mosaicing through Topology Inference and Local to Global Alignment, in: Proc. of European Conf. on Computer Vision, 1998, pp. 103 – 19.

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

[12] R. Pollefeys, R. Koch, L. Van Gool, Self-Calibration and Metric Reconstruction Inspite of Varying and Unknown Intrinsic Camera Parameters, Int. Journal of Computer Vision 32 (1) (1999) 7 – 25.

[13] U. Bhosle, Investigations in Image Mosaicing Under Different Problem Definitions, Ph.D. thesis, Department of Electrical Engineering, Indian Institute of Technology Bombay, India (2004).

[14] I. Zoghlami, R. Deriche, Using Geometric Corners to Build a 2-D Mosaic from a Set of Images, in: Proc. IEEE International Conference on Image Processing (ICIP), 1997, pp. 420 – 425.

[15] U. Bhosle, S. Chaudhuri, S. Dutta Roy, A Fast Method For Image Mosaicing Using Geometric Hashing, IETE Journal of Research: Special Issue On Visual Media Processing (2002) 317 – 324.

[16] U. Bhosle, S. Chaudhuri, S. Dutta Roy, Background Mosaicing of Scenes With Moving Objects, in: Proc. National Conference of Communication (NCC), 2003, pp. 84 – 89.

[17] C. Rothwell, Recognition using Projective Invariance, Ph.D. thesis, University of Oxford (1993).