Stereo Rectification of Un-calibrated and Heterogeneous Image-Pairs

S. Kumar*, C. Micheloni, C. Piciarelli and G.L. Foresti

Department of Mathematics and Computer Science, University of Udine, Via Della Scienze-206, Udine-33100, Italy

.

Abstract

In this paper, an algorithm to rectify a pair of heterogeneous images is presented. In particular, the pair of images is captured by using different sensors like static and dynamic cameras thus having different focal lengths and/or image resolutions. This combination of cameras is used as a binocular vision system. The pairs of images are made homogeneous based on the focal ratio of the two cameras. Zero padding and image shrinking operations are performed on stereo images to make them homogeneous in terms of size. The rectification transformations are calculated by solving a nonlinear constrained optimization problem for given pairs of corresponding points between stereo images. These pairs of matching points between stereo images are obtained using scale invariant features matching (SIFT matching). Experiments are performed to evaluate the performance of the proposed method using pairs of real stereo images. An error analysis is proposed to assess the improvements of the proposed method over the direct rectification. The proposed method is useful in stereoscopic as well as in video-surveillance applications.

Key words: Epipolar Geometry, Focal Ratio, Fundamental Matrix, SIFT Matching, Stereo Rectification, Zero Padding.

* **Corresponding Author**: S. Kumar
Email: *malikdma@gmail.com*
Dept. of Mathematics and Computer Science, University of Udine, Via Della Scienze-206, Udine-33100, Italy,
Phone: (+39)0432 558423

Preprint submitted to Elsevier *8 July 2009*

1 Introduction

Stereo vision is used to recover 3-D shape information of a real world scene from its two or more images taken from different viewpoints. The correspondence problem is a major issue in the application of stereo imagery. The correspondence problem is defined as locating a pair of image pixels from two different images, where these two pixels are projections of the same scene element. Given a point in one image, its corresponding point must lie on an epipolar line in the other image. This relationship is well known as epipolar constraint (Faugeras, 1993). The epipolar geometry can be calculated from the extrinsic and intrinsic parameters of a pair of stereo cameras.

If the two cameras are placed side by side on a base line and having identical intrinsic parameters then the image pair acquired by these cameras are known as rectilinear pair of stereo images. In these images corresponding points must lie on a same horizontal scan line (generally parallel to x axis) and the stereo matching problem reduces in a 1-D search along corresponding epipolar line instead of 2-D area search. When the epipolar geometry is not in this ideal form, the image pairs can be warped to make corresponding points lie on the same scan lines. This process is known as image rectification, and can be accomplished by applying 2D projective transforms, or homographies, to each image. The homography is a linear one to one transformation of the projective plane, which is represented by a 3×3 non-singular matrix.

The idea of rectification has long been used in photogrammetry as demonstrated in (Slama, 1980). The techniques originally used were optical-based, but now they are replaced by software methods that model the geometry of optical projection. In (Fusiello *et al.*, 2000), a software-based photogrammetric approach has been proposed which are similar to most of the computer vision ones when the knowledge of projection matrices or camera parameters is assumed. The similar techniques are demonstrated in (Ayache and Hansen, 1988,A). These methods require the knowledge of camera parameters to compute a pair of homographies. The necessity of camera calibration is one of their disadvantage. To overcome this disadvantage, several researchers have developed techniques, called projective rectification, to directly rectify images without using camera parameters. They utilized the epipolar geometry of the acquired images and various criteria to compute the homographies. In (Robert *et al.*, 1997), a method to find the best transformation that preserves orthogonality around image centers has been given. An approach using the minimization of the differences between matching points for the solution of homographies has been proposed in (Hartley, 1999). In (Loop and Zhang, 1999), a decomposition of each homography into projective and affine components is given.

Recently, a stereo rectification method which takes geometric distortion into account and tries to minimize the effects of resampling is given in (Gluckman and Nayar, 2001). A simple and efficient algorithm for generic two view stereo image rectification is presented by (Pollefeys *et al.*, 2001). The other available approaches include (Papadimitriou and Dennis, 1996) which considers only the special case of partially aligned cameras, and the approach proposed in (Al Shafan *et al.*, 2000) with a necessary requirement of epipolar geometry estimation. Although these proposed methods provided many possibilities for projective rectification, they all solve the problem indirectly. That is, they must explicitly estimate the fundamental matrix before rectification. Since the solution of fundamental matrix has its own uncertainty, this indirect approach might obtain unpredictable rectifying results (Zhang, 1998). In (Isgro and Trucco, 1999), a different procedure has been proposed by obtaining the homographies directly without first computing the fundamental matrix. However, in order to solve the problem of uniqueness in rectification, their method requires disparity minimization along the x-axis to generate a unique solution. In certain applications, this modification of x-axis disparity in rectification might be harmless; however, in applications where original x-axis disparity must be maintained (e.g. for stereoscopic viewing purpose), this constraint will make the algorithm useless. Moreover, the enforcement of minimizing x-axis disparity to obtain a single solution sometimes greatly distorts the image.

In (Wu and Yu, 2005), an improved algorithm for rectification of images by minimizing the distortion using a properly chosen shear transform is presented. However, this method is unable to perform well if the cameras are placed very far from each other. A method for the uncalibrated epipolar rectification which approximates the calibrated case by enforcing the rectifying transformation to be collinear induced by plane at infinity is presented in (Fusiello and Irsara, 2006). The advantage of their method is that there is no need of initial guess during minimization process. In (Zaharani *et al.*, 2004), a direct algorithm for the rectification of stereo images based on the estimation of homographies directly from geometric relationships has been given.

All the existing methods for stereo rectification in literature assume that the input images are homogeneous, i.e., camera pair images have same focal length and image resolution. In some video surveillance systems, a combination of static and dynamic cameras (e.g. Pan-Tilt-Zoom PTZ) are used to focus a target. The images captured from this combination of cameras are not homogeneous in terms of focal lengths, resolutions and illumination conditions. In this paper, a novel method for the rectification of a pair of non homogenous images is presented. To achieve such a result, the focal ratio between the focal lengths of the two images is computed for resizing the narrower image. The resize image has homogeneous focal information with respect to the wider image. To make homogenous even the image resolution a zero padding operation around the resize image is performed. Once these two steps have made the two images homogeneous the rectification process is run. Scale Invariant features are detected from both images to obtain the pairs of matching points. Rectifying transformations are obtained by solving a nonlinear constrained minimization problem. Experiments are conducted to rectify the various stereo pairs of real images.

This paper is organized as follows: In section 2, preliminaries of rectification process is given. Section 3 is devoted to detail description of presented methodology. In section 4, experimental results using our methodology are given and finally in section 5, the concluding remarks are given.

2 Preliminaries

The entire process of image rectification is performed in 2-D space. Therefore, the image planes are considered as projective planes in this work.

2.1 Epipolar Geometry

Suppose that I and I' are the images of a common scene and let m be a point in I. The locus of all points in R^3 that have the same projection m in I consists of a straight line through the center of the first camera. The projection of this entire line into the second camera is known as the epipolar line corresponding the point m. Any point m' in the second image plane which is the matching point of m must lie on this epipolar line. The epipolar lines in the second image plane corresponding to all points m_i of the first image intersect on a point e', called the epipole. The epipole e' is the converging point of the projections of the first image into the second image plane. Similarly, there is an epipole e in the first image defined by reversing the roles of the two image planes in the earlier discussion. Hence, the epipolar constrained is defined as

$$m_i'^T \; F \; m_i = 0, \qquad \text{for all} \;\; i \tag{1}$$

where F is called fundamental matrix. The matrix F is a 3×3 matrix with rank 2 that maps pixels from image I to image I' and vice versa. If m is a point in I then $Fm = l'$ is an epipolar line in I' corresponding to the point m. The epipoles are related by fundamental matrix as follows:

$$Fe = 0 = F^T e' \tag{2}$$

The fundamental matrix between a pair of images can be estimated directly from their extrinsic camera parameters. Some more algorithms are presented to estimate the fundamental matrix when extrinsic parameters are unknown like famous eight point algorithms.

2.2 Epipolar Geometry after Image Rectification

The rectification can be done by applying a transformation on the images, which maps for both images the epipole to a point at infinity. Therefore, the epipoles for a rectified pair of images are given as

$$e_\infty = e'_\infty = [1 \;\; 0 \;\; 0]^T, \tag{3}$$

and the fundamental matrix has the form

$$F_\infty = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \tag{4}$$

Let $(\overline{m}, \overline{m}')$ be a pair of corresponding points in the rectified images corresponding to the pair of points (m, m') in the original images. The epipolar constraints after rectification can be written as

$$\overline{m}^T F_\infty \overline{m}' = 0 \tag{5}$$

the rectification can be achieved by the following operations:

$$\overline{m} = Hm, \quad \overline{m}' = H'm' \tag{6}$$

where H and H' are the rectifying transformations for the left and right images, respectively. From (5) and (6), the following expression can be obtained

$$m'^T H'^T F_\infty H m = 0 \tag{7}$$

The homographies or transformations H and H' can be obtained by minimizing (7) for given several correspondence pairs (m, m'). In this paper, these pairs of correspondence points are obtained using the scale invariant features (SIFT) matching in the images I and I'.

3 Proposed Rectification Algorithm

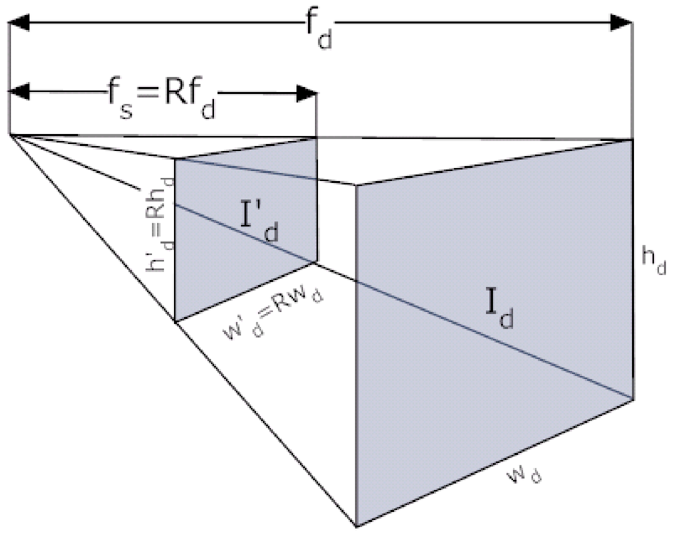

Fig. 1. Shrinking of image using the focal ratio information

3.1 Zoom-compensation

The aim of the proposed work is the rectification of a pair of images having heterogeneous internal image parameters such as zoom, resolution and size. If an image I_s of size $w_s \times h_s$ is captured with a static camera and the image I_d of size $w_d \times h_d$ is captured with a Pan-Tilt-Zoom (PTZ) camera on different zoom levels, then these two images have different focal lengths as well as resolutions. For a pin-hole camera model, the image formation equation can be given as

$$x_1 = f_1 \frac{X}{Z}, \qquad y_1 = f_1 \frac{Y}{Z} \tag{8}$$

where (x, y) denotes the image coordinates of a point P corresponding to its 3-D real world coordinate (X, Y, Z) and f_1 denotes the focal length of the camera on a certain zoom t_1. Let this zoom is changed (as in PTZ camera) to z_2 for any specific requirement of the surveillance system. Then the image formation equation can be written as

$$x_2 = f_2 \frac{X}{Z}, \qquad y_2 = f_2 \frac{Y}{Z} \tag{9}$$

where f_2 is the focal length corresponding to zoom t_2. The relation between the image pixels (x_1, y_1) and (x_2, y_2) can be obtained from above two equations as

$$x_1 = \frac{f_1}{f_2} x_2, \qquad y_1 = \frac{f_1}{f_2} y_2 \tag{10}$$

In this way the effect of unequal zoom levels can be compensated by focal ratio in perspective geometry (see Fig. 1). If we perform the rectification process directly on these images, then it would result in a large rectification error. To overcome this difficulty, the pair of images is made homogeneous before performing the rectification. The process for making the pair of images as homogeneous is given in the following steps.

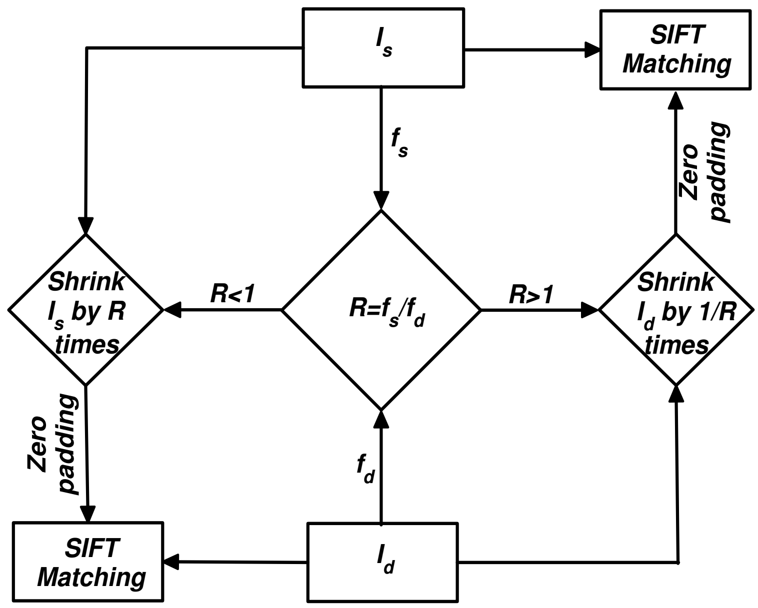

Fig. 2. Process for compensating the effect of unequal zoom

(1) Calculate the focal ratio $R = \frac{f_s}{f_d}$, where f_s and f_d are the focal lengths of the static and PTZ cameras, respectively. The focal length for static camera is computed once before use while the focal length for PTZ camera is estimated from the zoom information (see next sub-section).

(2) Check whether
- $R > 1$, then shrink the image I_d by $\frac{1}{R}$ times, let the shrunk image be I'_d of size $w'_d \times h'_d$. Go to step 3.
- $R < 1$, then shrink the image I_s by R times, let the shrunk image be I'_s of size $w'_s \times h'_s$. Go to step 4.
- $R = 1$, the images I_s and I_d are homogeneous, Go to step ccccc.

(3) Perform the zero padding of size $w'_d \times v_1$ on both horizontal sides of image I'_d, let the new image be I''_d. Also, perform the zero padding of size $u_1 \times h_d$

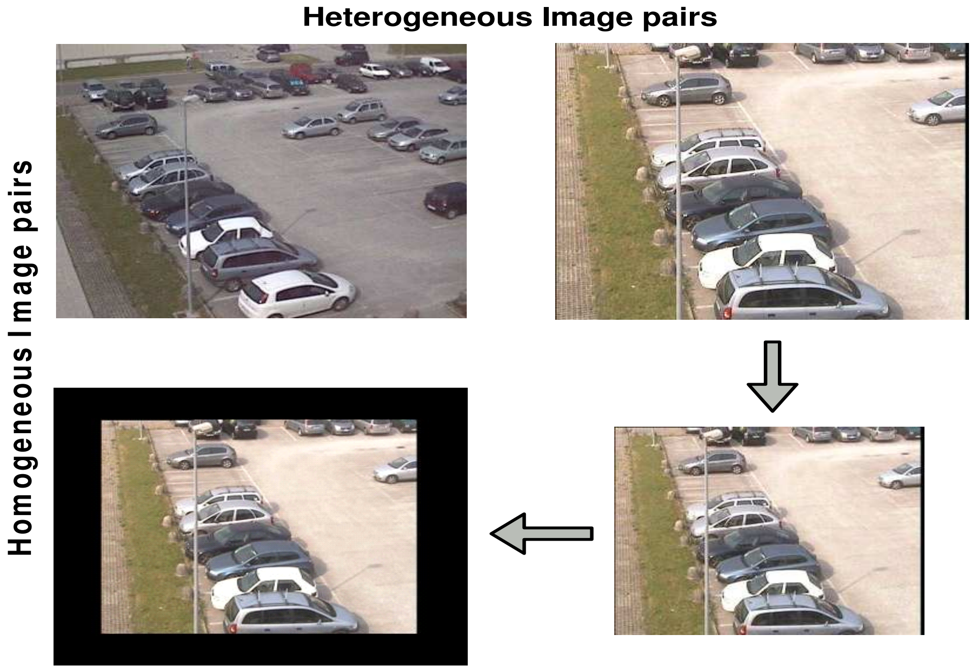

Fig. 3. Conversion into homogeneous pair from a heterogeneous pair of stereo images by compensating the unequal zoom effect when $f_s > f_d$.

on the both vertical sides of image I''_d. Where u_1 and v_1 can be given as

$$u_1 = \frac{|w_s - w'_d|}{2} \qquad \text{and} \qquad v_1 = \frac{|h_s - h'_d|}{2} \tag{11}$$

Let the new zero-padded image be I^h_d. Go to step 5.

(4) Perform the zero padding of size $w'_s \times v_2$ on both horizontal sides of image I'_s, let the new image be I''_d. Also, perform the zero padding of size $u_2 \times h_s$ on the both vertical sides of image I''_s. Where u_2 and v_2 can be given as

$$u_2 = \frac{|w_d - w'_s|}{2} \qquad \text{and} \qquad v_2 = \frac{|h_d - h'_s|}{2} \tag{12}$$

Let the new zero-padded image be I^h_s. Go to step 6.

(5) Check whether the size of images I_s and I^h_s is same or not. If it is not, go to step 3.

(6) Check whether the size of images I_d and I^h_d is same or not. If it is not, go to step 3.

This new pair of images is homogeneous in terms of both focal length as well as size. The overall process is shown in Fig. 2. Once the pair of homogeneous images is obtained, the next step is to obtain the pairs of matching points over this pair of images.

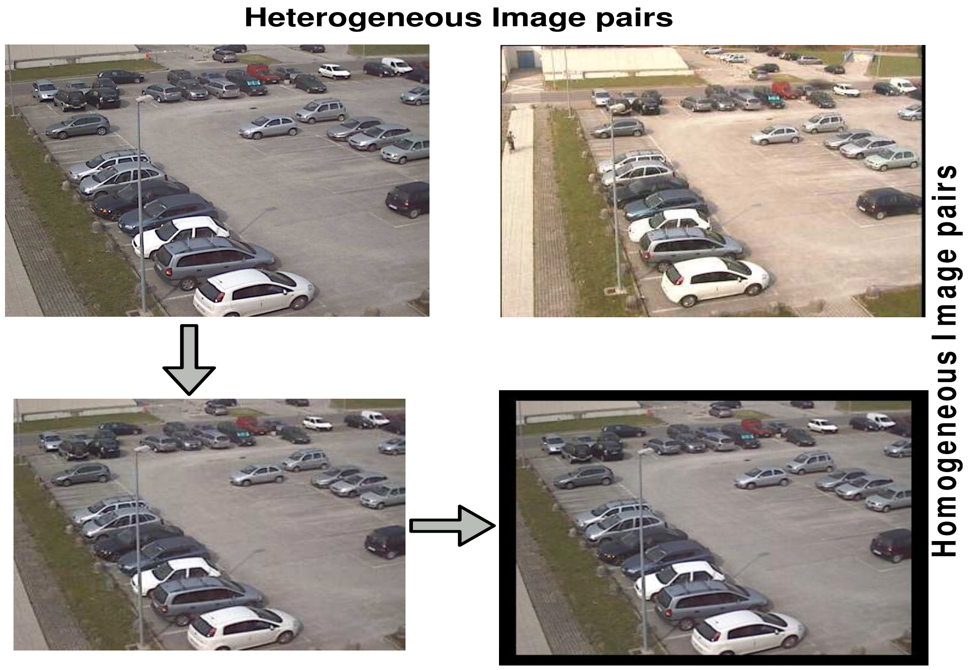

Fig. 4. Conversion into homogeneous pair from a heterogeneous pair of stereo images by compensating the unequal zoom effect when $f_s < f_d$.

3.2 Zoom to Focal Length Fitting

In the previous section, it has been described that the proposed algorithm is based on the zoom compensation process in case of heterogeneous image-pairs. The zoom is compensated by using a focal ratio information which required the focal lengths for the corresponding images. For a static camera, focal length can be estimated once in offline and use it by considering the fact that the image parameters (specifically focal length) will remain constant in whole process. However, the focal length is changed corresponding to change in zoom level in case of PTZ camera. Therefore, it is not easy to acquire the accurate focal length for a camera during the grabbing of images in real time when the system requires different zoom settings. Although, if it is not exact enough then it can affect the rectification accuracy in proposed algorithm. Here, the focal length is acquired in two steps, i.e., offline fitting of focal lengths corresponding to zoom settings and online estimation of it for a given particular zoom level.

Our aim is to find out a mapping between the zoom setting and focal length using several estimated focal lengths corresponding to their zoom settings. The whole range of zoom is sampled and for a particular zoom setting, estimation of the focal length is performed by taking two images at fixed pan and different tilt settings. Six different corresponding points are obtained between these two images using SIFT matching. The focal length is estimated according to a methodology given in (Li, 2006) using the six corresponding points. It is not practical to use this method in real system as it requires the images on two different pan/tilt settings by keeping the zoom constant. Therefore, we first estimate the focal length for each sampled zoom level for finding the mapping between zoom and focal length. A model have been proposed in (Trajkovic, 2002) for motorized lenses between a given zoom t and estimated focal length f as:

$$f(t) = \frac{a_0}{1 + a_1 t + a_2 t^2 + a_3 t^3 + \ldots + a_n t^n} \tag{13}$$

where the optimal value for the order n of the polynomial in the denominator can be computed by finding the ratio between focal lengths corresponding to minimum and maximum zoom using (13) for different n and compare it with the zoom power given by the manufacturer. It is found that $n = 2$ in this case. Coefficient a_0, a_1 and a_2 can be estimated by minimizing the following nonlinear function

$$C(a) = \sum_{i=1}^{n}((f(t_i) - \frac{a_0}{1 + a_1 t + a_2 t^2}))^2 \tag{14}$$

however from (14), the estimation of focal length is not reliable for small values of zoom, therefore (14) can be written as

$$C(b) = \sum_{i=1}^{n}(p(t_i) - (b_0 + b_1 t + b_2 t^2))^2 \tag{15}$$

where $p(t) = 1/f(t)$ denotes the lens power, $b_0 = 1/a_0$, $b_1 = a_1/a_0$ and $b_2 = a_2/a_0$. The minimization of (15) is reliable for lower as well as higher zoom settings. Total 26 zoom-ticks is considered by sampling the whole zoom range into 25 equal parts. For every zoom-tick a pair of images is captured by changing the pan setting at a constant tilt setting. The corresponding focal length is calculated and used in the minimization of (15). The values of b_0, b_1 and b_2 are stored corresponding to the minimum value of (15). Later, the focal length f for any given zoom level t can be found as

$$f(t) = \frac{a_0}{1 + a_1 t + a_2 t^2} \tag{16}$$

A graphical representation of the actual and the estimated focal lengths is given in Fig. 5. The mean of the error in estimated focal length is found 0.18 mm which is acceptable in proposed zoom compensation algorithm. Thus, the above method is reliable for estimating the focal length corresponding to a given zoom setting.

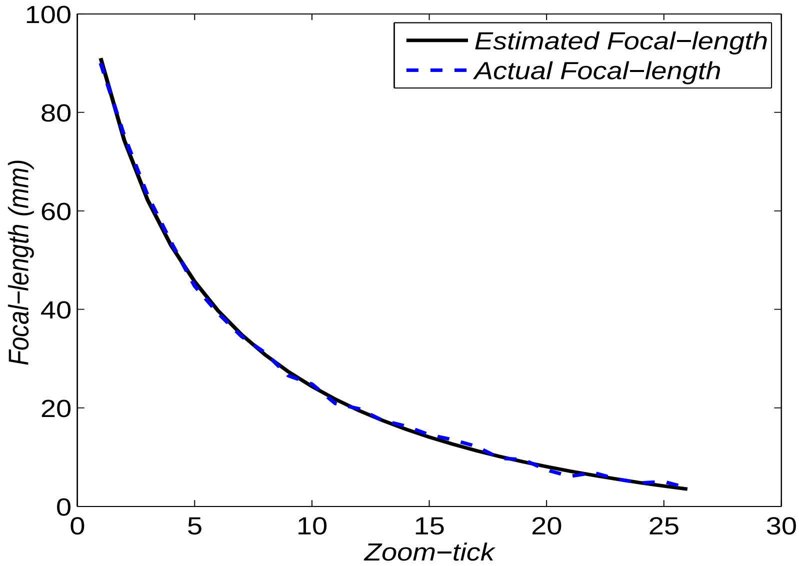

Fig. 5. Behavior of actual and estimated focal lengths corresponding to various zoom ticks for an AXIS 213 PTZ network camera.

3.3 SIFT Matching

The process to obtain the matching points from the pair of stereo images is divided into two steps. First, we detect the scale invariant features in each image separately. In the next step, matching process of these features is performed between stereo pair of images.

The process to identify locations in image scale space that are invariant with respect to image translation, scaling and rotation is based on the localization of a key. This task can be performed in following steps:

(1) Perform the convolution operation on input image I with the Gaussian function with variance $\sigma = \sqrt{2}$. Let this operation gives an image I_1.
(2) Repeat the step 1 on image I_1 to get a new image I_2.
(3) Subtract image I_2 from image I_1 to obtain the difference of Gaussian function as $\sqrt{2}$.
(4) Resample the image I_2 using bilinear interpolation with a pixel spacing of 1.5 in each direction. A 1.5 spacing means that each new sample will be a constant linear combination of 4-adjacent pixels. From this we generate a new pyramid level.
(5) Determine the maxima and minima of this scale-space function by comparing each pixel in the pyramid to its neighbors.
(6) Select key locations at maxima and minima of a difference of Gaussian function applied in scale space.

The scale invariant features can be detected from the locations of these keys.

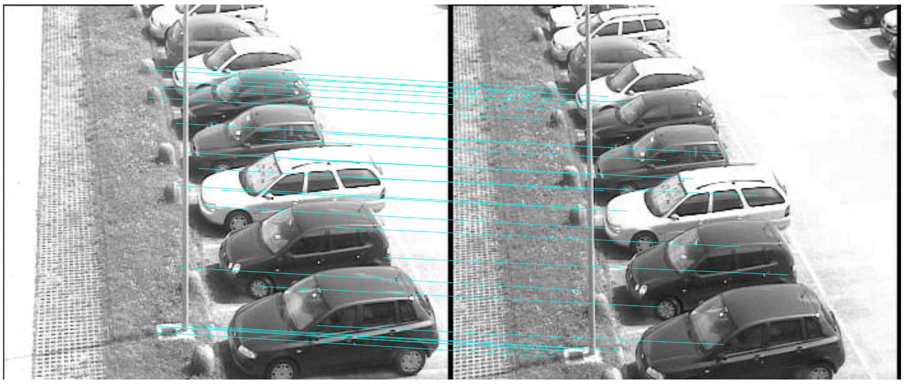

Fig. 6. An example of SIFT matching in pair of stereo images.

These features are detected on the exact key locations as well as locations around it so that reliable matching between different views of an object or scene can be performed. These features are invariant to not only image orientation but also image scale, and provide robust matching across a substantial range of affine distortion, change in 3-D viewpoint, addition of noise, and change in illumination.

For stereo image matching, SIFT features are extracted from left image and stored in a database. The right image features are matched by individually comparing each feature to this database and finding candidate matching features based on Euclidean distance of their feature vectors. We have performed features matching between stereo pair using the process given in (Lowe, 2004).

3.4 Image Rectification

In order to estimate the two homographies H and H', the pairs of corresponding points are used as input in the rectification process. Assume that the SIFT matching returns a number N of corresponding points in the two images, i.e., $(m_i, m_i'), \quad i = 1, 2, ..., N$. The two homographies can be computed by minimizing the following cost function

$$E(H, H') = \sum_{i=1}^{N} (m_i'^T H'^T F_\infty H m_i)^2 \tag{17}$$

As we know that the first row of F_∞ is a null vector this indicates that the equation 10 involves only the second and third rows of H and H'. However, rectification process is performed to simplify a stereo matching procedure, and if first row of H and H' is not chosen carefully in minimization, it may lead to a larger error in rectification process, hence to the failure in matching. Therefore, it is necessary to introduce some constraints in order to determine the first row of H. One constraint is the minimization of the vertical distances between corresponding pairs, i.e., distance between corresponding epipolar lines along vertical axis

$$d(H, H') = \sum_{i=1}^{N} (|(Hm_i)_y - (H'm'_i)_y|) \tag{18}$$

where $(.)_y$ indicates y coordinates. In other words we want to minimize the vertical distance between corresponding points. The nonlinear objective function of optimization problem together with constraints can be defined as

$$F(H, H') = E(H, H') + \lambda d(H, H') \tag{19}$$

where $\lambda = [\lambda_1,\ \lambda_2,\ \lambda_3]$ is a row vector. The minimization of (12) is a nonlinear minimization problem. To perform this minimization we choose the Levenberg-Marquardt algorithm because of its effectiveness and popularity. Before applying this process, we need to derive the Jacobian matrix of (12), i.e., $\frac{\partial F}{\partial \phi}$, where the vector ϕ contains all unknown parameters of H, H' and λ. A good initial estimate is important to guarantee the convergence to a correct solution. To overcome this we use local search, i.e., first rectifies subsampled images, where the difference between coordinates is comparatively small. Later, the estimated homographies are used for the subsampled level to initialize the estimation for the finer levels.

4 Results and Discussions

To evaluate the performance of the proposed rectification algorithm, experiments have been performed on the images from a parking lot. An AXIS-213 PTZ network camera and a static cameras placed at the roof of the University building are used to grab these sequences. The images captured by these cameras are both of size 352×288. The zoom settings of these cameras have been acquired online and the focal information are estimated from these zoom settings. Our aim behind to propose this algorithm is obtained the more accuracy in rectification for heterogeneous stereo images, i.e., when stereo images have different internal image parameters. Therefore, three different pairs of images having focal ratio 1.0, 0.85 and 0.70 have are used for experimental study. The SIFT matching is performed to obtain the pairs of corresponding points in the left and right images. The algorithm has been tasted for a fixed number of matching points for all pairs of images.

Fig. 7. Rectification of pair of stereo images having focal ratio 1.0. First row: pair of stereo images, Second row: rectified pair of stereo images.

Several criterions can be used for computing the error in a rectified pair of images. The more general criterion is the mean of the vertical distances between the corresponding points, i.e., the distance along the vertical axis of the corresponding points or epipolar lines in rectified pairs.

$$r_i = |(Hm_i)_y - (H'm'_i)_y|$$

Since, rectification gives a pair of images in which corresponding epipolar lines should be collinear and parallel to horizontal axis, therefore the above criterion gives more sense for computing the error in a rectified pair of images. The given results aim to show the improved performance of the proposed solution with respect to a method that rectifies images directly. An exception is made to resize images to make them homogenous in size. In the reminder of this section we will refer to this last technique with the terms "'*direct rectification*"'.

The first set of experiments consists in computing the error in rectification when the pair of images has been obtained with identical focal lengths. In this context, Fig. 7 presents the rectification results for a pair of images having focal ratio 1.0. The error in rectified pair of images is 0.14 when 15 pairs of matching points are used. It implies that the algorithm works very well for the homogeneous images. In particular, it is worth noticing that in case of homogenous images the proposed solution shows the same error of the direct rectification (see Fig. 12). This is due to the fact that in this case the proposed

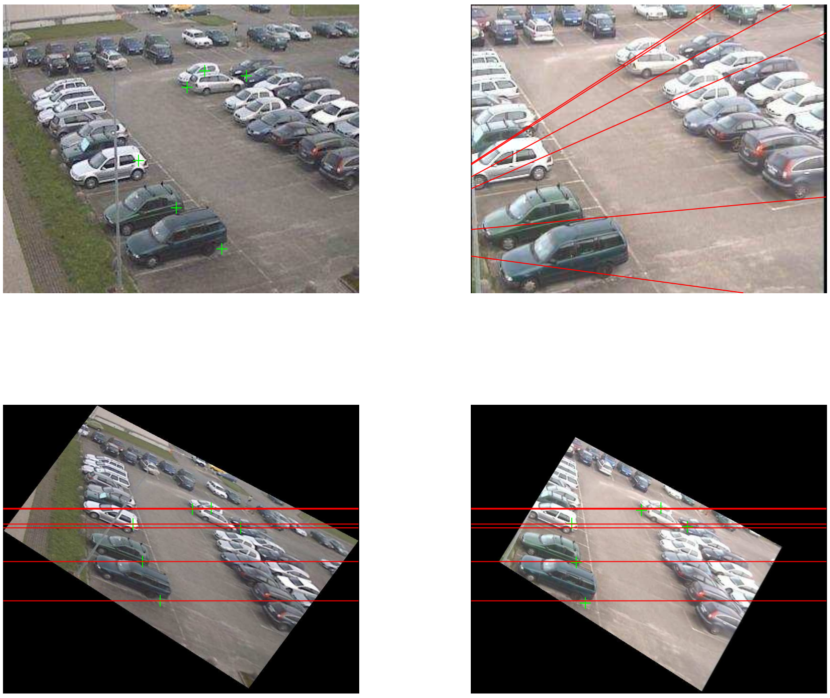

Fig. 8. Direct rectification of pair of stereo images having focal ratio 0.85. First row: pair of heterogeneous stereo images, Second row: rectified pair of stereo images.

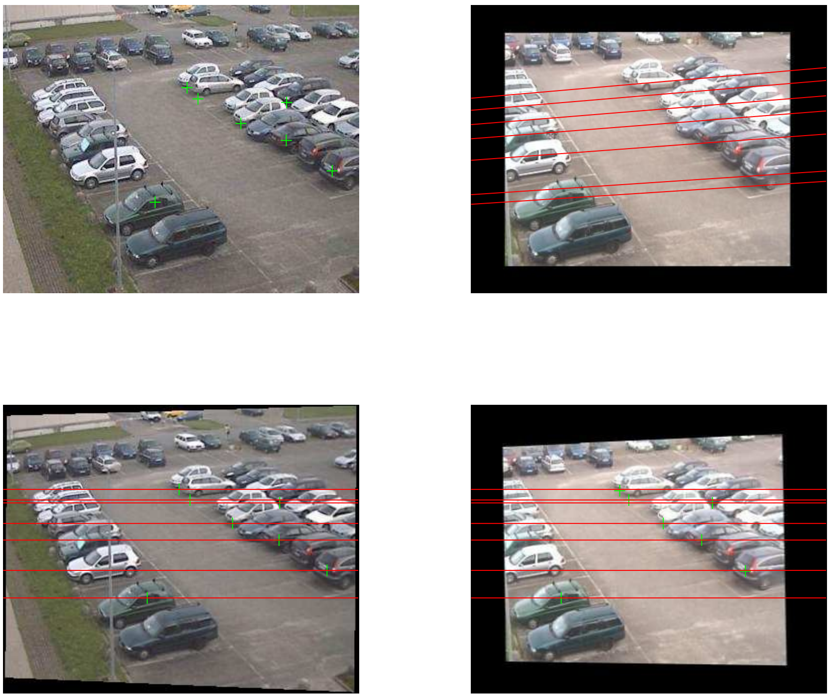

Fig. 9. Proposed rectification of pair of stereo images having focal ratio 0.85. First row: pair of homogeneous stereo images, Second row: pair of homogeneous stereo images.

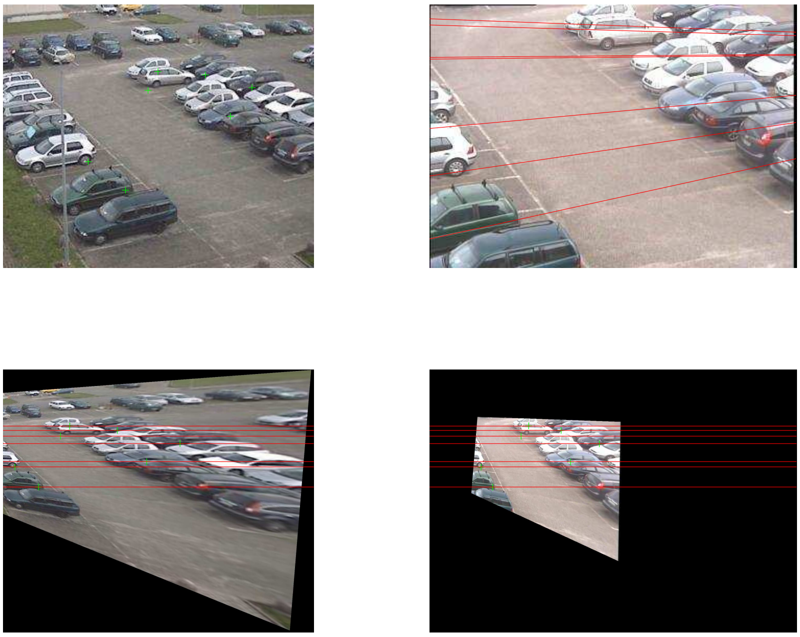

Fig. 10. Direct rectification of pair of stereo images having focal ratio 0.70. First row: pair of stereo images (heterogeneous), Second row: rectified pair of stereo images.

Fig. 11. Direct rectification of pair of stereo images having focal ratio 0.70. First row: pair of stereo images (homogeneous), Second row: pair of homogeneous stereo images.

solution makes the same operations of a direct rectification technique.

Since the main goal of the proposed algorithm is the rectification of non-homogeneous images, we run a set of experiments by progressively reducing the focal ratio. In Fig. 8 and Fig.9, the rectification results are shown for a pair of images having focal ration 0.85 and tasted for 15 pairs of matching points. Fig. 8 shows the rectified results obtained by employing the direct rectification process while the rectified results obtained with the proposed solution that brings to rectification homogeneous pair of images is shown in Fig. 9. As can be seen from these two sets of images the results obtained by the proposed solution start to be visible. In addition if we look for the performance comparison at Fig. 9 shows how the proposed solution performs much better than the direct approach. The same comparison is carried out with Fig. 10 and Fig. 11, the rectification results has been shown for a pair of images having focal ratio 0.70 and tested on 15 pairs of matching points. Fig. 10 shows the rectified results when pair is non-homogeneous while the rectified results for an homogeneous pair of images is shown in Fig. 11.

A graphically comparison between both approaches, i.e., when images have been rectified using a direct rectification process and the proposed rectification process (after making the two images homogeneous in terms of focal length and size) has been shown by Fig. 12. It represents that the error is very high when the rectification process has been performed directly without using homogeneous image pair while the error is very small when using the proposed algorithm up to a focal ratio 0.50. When the focal ratio is reduced less than 0.50 then a very high gradually increment is noticed in the rectification error for both the approaches. It is not easy to performed SIFT matching in the cases when focal ratio is less than 0.50 is the other limitation of the proposed algorithm.

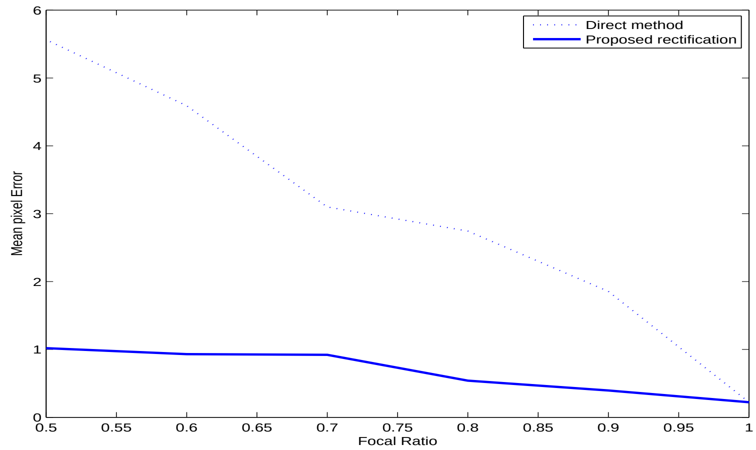

Fig. 12. Error in direct and proposed rectification process corresponding to different focal ratios.

5 Conclusions

We have presented an approach for the rectification of uncalibrated and non-homogeneous pairs of stereo images. The pair of images has been made homogeneous in terms of focal lengths and sizes. This is based on a focal ratio information and then by performing zero padding on the shrunk image. The pairs of corresponding points have been obtained using SIFT matching in stereo pair of images. The rectification transformations have been obtained by solving a nonlinear optimization problem. A constraint has been added in the objective function of nonlinear minimization problem to improve the accuracy of proposed algorithm. The algorithm has been tested on the real images acquired by a combination of static and PTZ cameras. From the experimental results, it is concluded that the algorithm is highly accurate for any kind of stereo imaging systems with minimal rectification and distortion error. This method can be used in video surveillance applications as well as to estimate the disparity in stereo images.

Acknowledgements

This work was partially supported by the Italian Ministry of University and Scientific Research within the framework of the project entitled *Ambient Intelligence: event analysis, sensor reconfiguration and multimodal interfaces* (2006-2008). S. Kumar is also thankful to Department of Mathematics and Computer Science, University of Udine for the financial support during this work under the grant MIUR n. 179 dd. 29/01/2007 provided by Italian Ministry of University and Scientific Research.

References

Al-Shalfan, K., Haigh, J., Ipson, S., 2000. Direct algorithm for rectifying pairs of uncalibrated images, Electronics Letters 36 (5), 419–420.

Ayache, N., Hansen, C., 1988. Rectification of images for binocular and trinocular stereovision, in: IEEE international Conference on Pattern Recogination, Rome, Italy.

Ayache, N., Lustman, F., 1991. Trinocular stereo vision for robotics, IEEE Trans. on Pattern Analysis and Machine Intelligence 13, 73–85.

Faugeras, O., 1993. Three-Dimensional Computer Vision, MIT Press, Cambridge, MA, USA.

Fusiello, A., Irsara, L., 2006. Quasi-euclidean uncalibrated epipolar rectification, Research Report RR 43/2006, Department of Computer Scince, Univ. of Verona.

Fusiello, A., Trucco, E., Verri, A., 2000. A compact algorithm for rectification of stereo pairs, Machine Vision and Applications 12, 16–22.

Gluckman, J., Nayar, S., 2001. Rectifying transformations that minimize resampling effects, in: IEEE International Conference on Computer Vision and Pattern Recogination, Kauai, HI, USA.

Hartley, R., 1999. Theory and practice of projective rectification, Intrnational Journal of Computer Vision, 35, 115–127.

Isgro, F., Trucco, E., 1999. On robust rectification for uncalibrated images, in: IEEE International Conference on Image Analysis and Processing.

Li, H., 2006. A Simple Solution to the Six-Point Two-View Focal- Length Problem, Multiview Geometry and 3D Methods (LNCS).'

Loop, C., Zhang, Z., 1999. Computing rectifying homographies for stereo vision, in: IEEE International Conference on Computer Vision and Pattern Recogination, Fort Collins, CO, USA.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2), 91–110.

Papadimitriou, D., Dennis, T., 1996. Epipolar line estimation and rectification for stereo image pairs, IEEE Transactions on Image Processing 5 (4), 672–676.

Pollefeys, M., Kock, R., Gool, L., 2001. A simple and efficient rectification method for general motion, in: IEEE International conference on Image Processing, Corfu, Greece.

Robert, L., Zeller, C., Faugeras, O., 1997. Applications of non-metric vision to some visually-guided robotics tasks, In : Visual Navigation: From Biological Systems to Unmanned Ground Vehicles, vol. 2, Lawrence Erlbaum Associates.

Slama, C., 1980. Manual of Photogrammetry, American Society of Photogrammetry.

Trajkovic M., 2002.Interactive Calibration of a PTZ Camera for Surveillance Applications, Asian Conference on Computer Vision.

Wu, H., Yu, Y.-H., 2005. Projective rectification with reduced geometrical distortion for stereo vision and stereoscopic video, Journal of Intelligent and Robotics Systems, 42,71–94.

Zahrani, A., Ipson, S., Haigh, J., 2004. Application of a direct algorithm for the rectification of uncalbrated images, Information Sciences, 160, 53–71.

Zhang, Z., 1998. Determining the epipolar geometry and its uncertainty: A review, International Journal of Computer Vision, 27, 161–195.