Pattern Recognition Letters 31 (2010) 1445–1452

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Stereo rectification of uncalibrated and heterogeneous images

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

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

ARTICLE INFO

Article history:
Received 18 July 2008
Received in revised form 28 October 2009
Available online 27 March 2010
Communicated by B.C. Lovell

Keywords:
Epipolar geometry
Focal ratio
Fundamental matrix
SIFT matching
Stereo rectification
Zoom refinement

ABSTRACT

In this paper, an algorithm for rectifying heterogeneous and uncalibrated pairs of stereo images is presented. In particular, a pair of images is captured by using a combination of static and dynamic cameras at unequal zoom, thus having different focal lengths and/or image resolutions. The rectification of such pairs of images is made in two steps. In the first step, image shrinking based on focal ratios is performed for compensating the effect of unequal zoom levels followed by a zero padding on the smaller image for making the images of equal size. In the second step, rectification transformations are calculated by solving a nonlinear constrained optimization problem for a given set of pairs of corresponding points (SIFT descriptors) between stereo images. Experiments are performed to evaluate the performance of the proposed method and assess the improvements of the proposed method over direct rectification.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Stereo vision is used to recover 3-D shape information of a real world scene from two or more images taken from different viewpoints. The correspondence problem is a major issue in the application of stereo imagery. It 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 known as epipolar constraint (Faugeras, 1993). If the two cameras are placed side by side on a base line and have identical intrinsic parameters then the acquired images are known as a rectified pair of stereo images. In these images, corresponding points must lie on a same horizontal scan line. 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 2-D projective transformations on each image. These transformations are linear, one to one mapping, and represented by a 3×3 non-singular matrix.

Initially, the idea of rectification has been used in photogrammetry as demonstrated in (Slama, 1980). The techniques originally were optical-based, but later replaced by software methods that model the geometry of optical projection. In (Fusiello et al., 2000), a rectification approach has been proposed by using the knowledge of known camera parameters. Similar techniques are demonstrated in (Ayache and Hansen, 1988; Ayache and Lustman, 1991). The necessity of known calibration parameters is one of the disadvantage of these methods. Projective rectification has been introduced to overcome this disadvantage, by using epipolar geometry with various constraints. In (Robert et al., 1997), a method to find the best transformation that preserves orthogonality around image centers has been given. An approach using difference minimisation between matching points for computing rectification transformations has been proposed in (Hartley, 1999). In (Loop and Zhang, 1999), rectification transformation has been obtained by decomposing it into the projective and affine components.

Recently, a stereo rectification method which takes geometric distortion into account and tries to minimize the effects of resampling has been given in (Gluckman and Nayar, 2001). A simple and efficient algorithm for generic two view stereo image rectification has been presented in (Pollefeys et al., 2001). The other available approaches include (Papadimitriou and Dennis, 1996) which considered only the special case of partially aligned cameras, and the approach proposed in (Al-Shalfan et al., 2000) with a necessary requirement of epipolar geometry estimation. All these methods compute projective rectifications in an indirect way, since an explicit estimation of fundamental matrix before rectification is needed. Since the computation 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 rectification transformations directly without computing the fundamental matrix with a disparity minimization based uniqueness criterion. However in some cases, the enforcement of minimizing x-axis disparity gives a distorted rectified images. A modification of this approach has been given in (Wu and Yu, 2005) using a proper shear

* Corresponding author. Tel./fax: +39 0432 558423.
E-mail address: sanjeev.kumar@dimi.uniud.it (S. Kumar).

0167-8655/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2010.03.019

transform. A method for uncalibrated epipolar rectification which approximates the calibrated case by enforcing the rectifying transformation to be collinear induced by plane at infinity has been presented in (Fusiello and Israra, 2008). The advantage of their method is that there is no need of initial guess during the minimization process. In (Zahrani et al., 2004), a direct algorithm for stereo images rectification has been given, based on the estimation of homographies directly from geometric relationships.

All the existing methods for stereo rectification assume that the input images are homogeneous, i.e., images are obtained at the same zoom levels and have the same size. In many surveillance-oriented applicative fields, however, other configurations can be used, such a combination of static and Pan-tilt-zoom (PTZ) cameras for close focusing on a target in large environment as a master–slave framework. The acquired images from these system are not homogeneous in terms of focal lengths, sizes and illumination conditions. A direct rectification of such heterogeneous images gives large rectification errors in the resulting images. Therefore, a compensation of this heterogeneity is required before rectification. This heterogeneity is involved due to unequal zoom levels, therefore a compensation can be made to avoid such heterogeneity by shrinking the higher resolution image (having narrow field of view) by the ratio of their focal lengths. In this way, the shrunk image has homogeneous focal information with respect to the other image but possibly has a different size. The sizes of these two images can be made equal by a zero padding around the smaller image. Finally, this homogeneous image pair is rectified by minimizing the epipolar constraints based objective function for given pairs of matching points.

In brief, the main contribution of this paper includes a modified stereo rectification approach in case of heterogeneous pairs of images. This modified approach contains a pre-rectification compensation process of unequal zoom and sizes between the images of a stereo pair. This compensation is based on the focal ratio of the corresponding pair of images by shrinking the high resolution image followed by a zero padding on smaller image. A required formulation is given to acquire this focal ratio information from zoom data for any pair of cameras. Experimental results are presented for representing the superiority of proposed approach over a direct rectification approach in case of heterogeneous pairs of images.

This paper is organized as follows: in Section 2, preliminaries of the rectification process are given. Section 3 is devoted to a detailed description of proposed two-steps rectification algorithm. In Section 4, experimental results are presented and finally Section 5 is devoted to concluding remarks.

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 two images of the same scene from different viewpoints, and let x be a point in I. The locus of all points in R^3 that have the same projection x 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 x. Any point x' in the second image plane I' that is the matching point of x must lie on this epipolar line. The epipolar lines in the second image plane corresponding to all the points x_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 constraint is defined as

$$x_i'^T F x_i = 0, \quad \text{for all } i \tag{1}$$

where F is a 3×3 with rank 2 matrix, called fundamental matrix that maps pixels from image I to image I' and vice versa. 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. Different algorithms have been proposed to estimate the fundamental matrix when extrinsic parameters are unknown.

2.2. Epipolar geometry after image rectification

The rectification can be achieved by computing a pair of transformation matrices that map the epipoles 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 $(\bar{x}, \bar{x}')$ be a pair of corresponding points in the rectified images corresponding to the pair of points (x, x') in the original images. The epipolar constraints after rectification can be written as

$$\bar{x}'^T F_\infty \bar{x} = 0 \tag{5}$$

the rectification can be achieved by the following operations:

$$\bar{x} = Hx, \quad \bar{x}' = H'x' \tag{6}$$

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:

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

The transformations H and H' can be obtained by minimizing (7) for given a set of correspondence points.

3. Proposed rectification algorithm

Two steps are required for rectifying a heterogeneous pair of stereo images. First, a pair of heterogeneous images is transformed into a homogeneous pair. In the second step, rectification transformations are found by minimizing the epipolar constraint based nonlinear function for a given set of corresponding points in the new homogeneous pair of images.

3.1. Zoom-compensation

For a pin-hole camera model, the image formation equation can be given as

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

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 level t_1. If the zoom t_1 is changed into t_2 for a specific requirement of the surveillance system, the image formation equation can be written as

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

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 the above two equations as

$$x_1 = Rx_2, \quad y_1 = Ry_2 \tag{10}$$

where $R = \frac{f_1}{f_2}$ is called focal ratio. 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 without compensating this effect, it would result in a large rectification and distortion error. Therefore, the heterogeneity compensation process is required for obtaining a better pair of rectified images.

Let a pair of heterogeneous images be I_s of size $w_s \times h_s$ and I_d of size $w_d \times h_d$ captured with static and PTZ cameras, respectively, with different zoom settings. The process for converting this pair into homogeneous is accomplished by using following two operations.

3.1.1. Image shrinking

This operation is performed for compensating the effect of unequal zoom settings between images of a stereo pair. It requires the following steps:

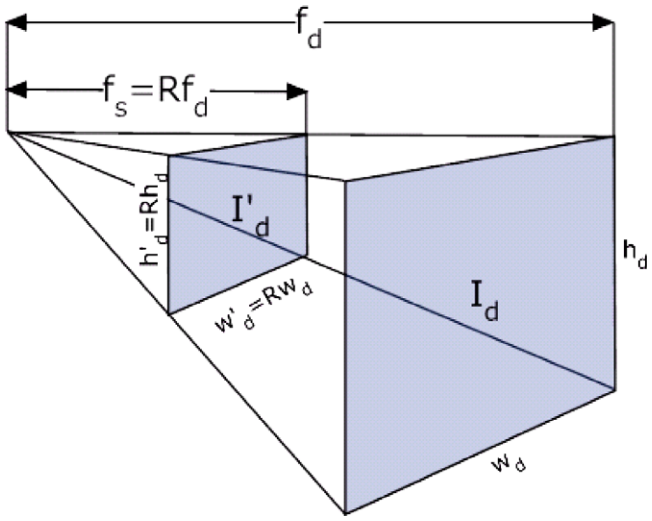

Fig. 1. Shrinking of image using the focal ratio information.

(1) Calculate the focal ratio $R = f_d/f_s$, where f_s and f_d are the focal lengths of the static and PTZ cameras, respectively.
(2) Check, if
   (a) $R > 1$, then shrink the image I_d by R times.
   (b) $R < 1$, then shrink the image I_s by $\frac{1}{R}$ times
   (c) $R = 1$, do not shrink any image.

It is obvious that the size of shrunk image is changed. For sake of simplicity, let $R > 1$ and the image I_d shrunk into I'_d of size $w'_d \times h'_d$. This new pair of images (I_s, I'_d) is homogeneous in terms of focal lengths.

3.1.2. Zero padding

The zero padding based operation is performed for making the images I_s and I'_d homogeneous in size. It requires the following steps:

(1) Find the smaller image in size from I_s and I'_d.
(2) If $I_s > I'_d$, perform zero padding of size $w'_d \times v_1$ on the horizontal sides of I'_d, followed by another zero padding of size $u_1 \times h_s$ on its vertical sides, where u_1 and v_1 can be given as

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

Let the new zero-padded image be I^h_d. In this way, (I_s, I^h_d) is the homogeneous pair of images.
(3) If $I_s < I'_d$, perform zero padding of size $w_s \times v_2$ on horizontal sides followed by another zero padding of size $u_2 \times h'_d$ on the vertical sides of image I_s. Where u_2 and v_2 can be given as

$$u_2 = \frac{|w'_d - w_s|}{2} \quad \text{and} \quad v_2 = \frac{|h'_d - h_s|}{2} \tag{12}$$

Let the new zero-padded image be I^h_s. In this way, (I^h_s, I'_d) is the homogeneous pair of images.
(4) If $I_s = I'_d$, zero padding is not required and (I_s, I'_d) is the homogeneous pair of images.

The pair of images obtained after the above mentioned operations will be homogeneous in terms of focal length as well as size. Fig. 2 represents a flow diagram for this heterogeneity compensation process. Fig. 3 illustrates the examples of the compensation process in case of $R > 1$ and $R < 1$.

3.2. Focal length estimation from zoom

In the previous section, it has been described that the proposed algorithm is based on the zoom compensation process in case of

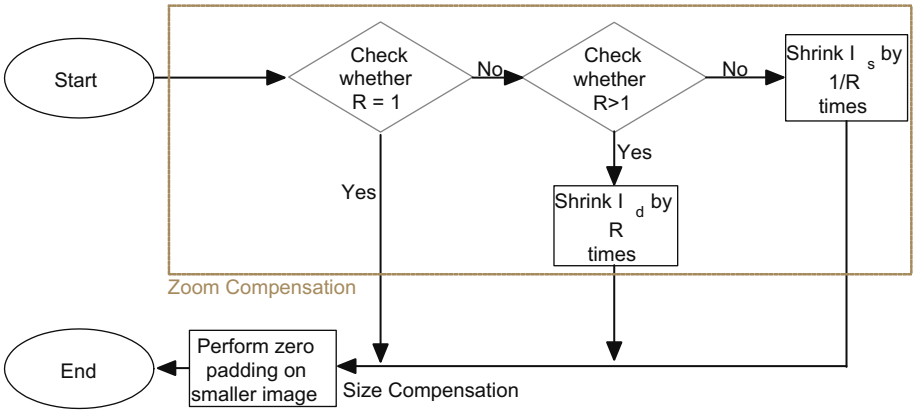

Fig. 2. Process for compensating the effect of heterogeneity.

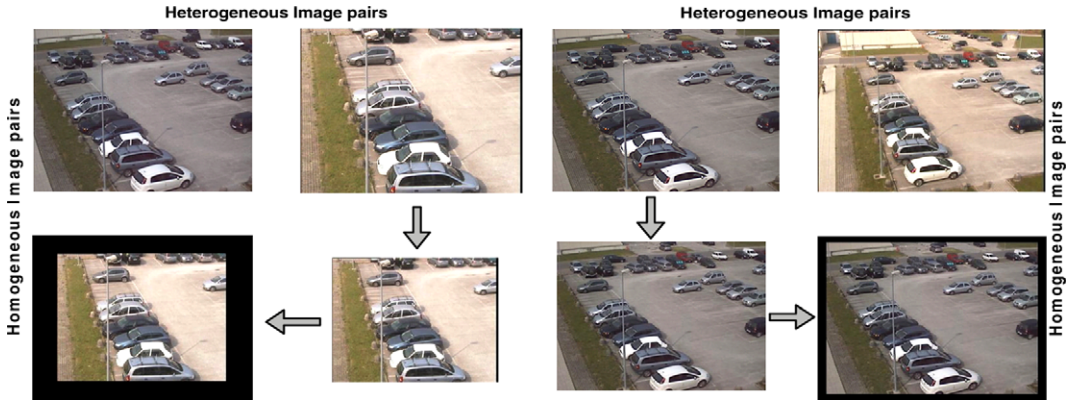

Fig. 3. Conversion into homogeneous pair of an heterogeneous pair of stereo images when zoom of PTZ camera is greater (left) and lower (right) than the static camera.

heterogeneous image-pairs. The zoom is compensated by using a focal ratio information which requires the focal lengths for the corresponding images. For a static camera, the focal length can be estimated offline and it will remain constant for the whole process. However, in the case of PTZ cameras, the focal length changes corresponding to variations in the zoom level. 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. Here, several focal lengths are acquired offline for different zoom settings, and online focal length estimation can be obtained through interpolation.

3.2.1. Zoom refinement

A PTZ camera can provide instantaneous information about the zoom parameter. However, this information contains a small error due to mainly two reasons: (1) mechanical clearance, and (2) lack of synchronization between image capturing and PTZ parameters acquiring, especially when the camera is moving. A refinement of this error can be done by considering several calibrated images from the same camera. Let $(I_1, I_2, \ldots, I_n)$ be n calibrated images captured at different zoom ticks $(t_1, t_2, \ldots, t_n)$, respectively. Denote x_1 as a feature point in I_1, and use SIFT descriptor (Lowe, 2004) to find the matching points of x_1 in the other images. Let denote the matching points as $\{x_i\}$ in a calibrated image I_i. Assume t_i^0 is the ground truth zoom parameter for image I_i, and $\Delta t_i (= t_i - t_i^0)$ is the error in zoom parameters for image I_i. Our aim is to calculate Δt_i for a PTZ camera.

Let X_i, X_0 and $\hat{x}_i$ be the estimated world coordinates, ground-truth world coordinates and homogeneous normalized image coordinates of x_i, respectively. According to Wan and Zhou (2008), the image formation equation for pin-hole camera model can be written as:

$$\hat{x}_i = \lambda K(t_i)\Re(\theta_i^1, \theta_i^2)X_i \tag{13}$$

where λ is the scaling factor and $(\theta_i^1, \theta_i^2, t_i)$ denotes a vector of pan, tilt and zoom parameters. The matrix $\Re(\theta^1, \theta^2)$ is the camera orientation matrix and $K(t) = diag[f(t), f(t), 1]$ is a scaling matrix dependent on zoom t (equivalent to intrinsic matrix with unit pixel aspect ratio, no radial distortion and principal point as zoom center). In a reverse way, (13) can be written as

$$X_i = \lambda P(\theta_i^1, \theta_i^2, t_i)\hat{x}_i, \quad \text{and} \quad X_0 = \lambda P(\theta_i^1, \theta_i^2, t_i^0)\hat{x}_i \tag{14}$$

where $P = \Re^{-1}K^{-1}$, as we are considering the changes only in zoom whereas pan and tilt parameters are kept constant. In this context, taking the first order Taylor expansion of X_i in (14) at t_i^0, we have

$$X_i \simeq \lambda P(\theta_i^1, \theta_i^2, t_i^0)\hat{x}_i + \left(\frac{\partial(\lambda P(\theta_i^1, \theta_i^2, t_i)\hat{x}_i)}{\partial t_i}\right)^T \Delta t_i \simeq X_0 + M_i \Delta t_i \tag{15}$$

where M_i is a 3×3 invertible matrix, so

$$\Delta t_i = M_i^{-1}(X_i - X_0) \tag{16}$$

Here the unknown X_0 cab be found by minimizing $E = \sum_i \|\Delta t_i\|$ as in (Wan and Zhou, 2009). From (16), this objective function can be written as

$$E = \sum_i (X_i - X_0)^T B_i (X_i - X_0) \tag{17}$$

where $B_i = (M_i^{-1})^T M_i^{-1}$. By setting the partial derivative of E with respect to X_0 equivalent to zero, we obtain

$$X_0 = \left(\sum_i (B_i)^{-1}\right)\left(\sum_i (B_i X_i)\right) \tag{18}$$

which can be regarded as the weighted average of X_i. Once the ground truth world coordinates X_0 are calculated using (18), we compute the Δt_i using (16) for I_i. The precision and robustness of Δt_i can be improved by taking several feature points x_1^j for an image, and applying the weighted least-square method in its calculation. Let m_j be the number of points matched with x_1^j in m different images, the larger m_j, the higher the weight. Finally the refined zoom parameter t_i^0 is obtained as $t_i^0 = t_i + \Delta t_i$, and it is used as the acquired zoom value for further focal length estimation process.

3.2.2. Focal length to zoom fitting

The whole zoom range of the employed PTZ camera is sampled into n ticks, and a pair of images is captured by tilting the camera at each sampled tick. This procedure is repeated to acquire the 10 samples of image pairs. In this way for a zoom tick, we acquired 10 pairs of images and total $n \times 10$ pairs of images for whole zoom range. We calibrate each pair to obtain the 10 samples of focal length for a zoom tick and calculate the mean value of these 10 samples and consider it as calibrated focal length corresponding to the zoom value of this tick. A technique proposed in (Trajkovic, 2002) is adopted to define a function linking a zoom tick to its corresponding focal length. In the case of motorized lenses, the relation between a given zoom tick t and corresponding focal length f is

$$f(t) = \frac{a_0}{1 + a_1 t + a_2 t^2 + a_3 t^3 + \ldots + a_n t^n} \tag{19}$$

where the order n is camera dependent. For the adopted camera, following the methodology in (Trajkovic, 2002), the estimated opti-

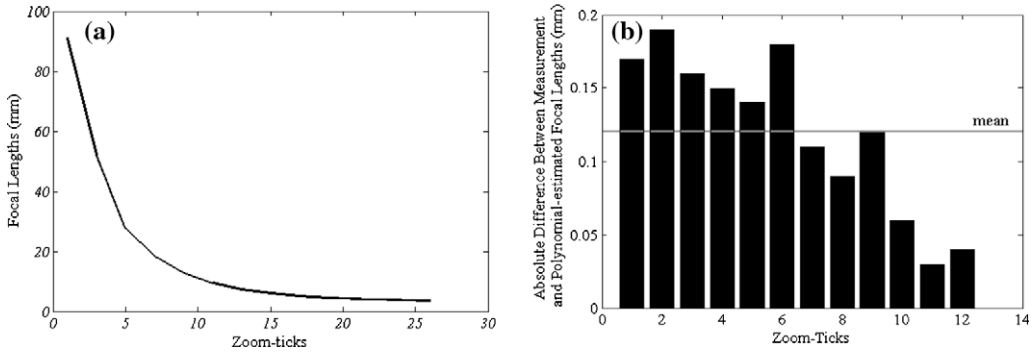

Fig. 4. Focal length estimation: (a) typical curve showing focal length as a function of zoom ticks; (b) absolute difference between calibrated and polynomial-estimated focal lengths for different 12 zoom ticks.

mal value of n is 2. Therefore, a_0, a_1 and a_2 can be estimated by minimizing the following nonlinear function:

$$C(a) = \sum_{i=1}^{n} \left(f(t_i) - \frac{a_0}{1 + a_1 t + a_2 t^2} \right)^2 \tag{20}$$

However, from (20), the estimation of the focal length is not reliable for small as well as higher values of zoom, therefore (20) can be written as

$$C(b) = \sum_{i=1}^{n} \left(p(t_i) - (b_0 + b_1 t_i + b_2 t_i^2) \right)^2 \tag{21}$$

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 (21) is reliable for lower as well as higher zoom settings. The whole zoom range is sampled into $n(= 25)$ equal parts for the adopted 26× zoom camera. The calibrated focal lengths for each zoom tick are obtained and used in the minimization of (21). The values b_0, b_1 and b_2 corresponding to the minimum value of $C(b)$ are chosen to acquire the optimal values of a_0, a_1 and a_2. In real time, the focal length f for any given zoom tick t_c is estimated as

$$f(t_c) = \frac{a_0}{(1 + a_1 t_c + a_2 t_c^2)} \tag{22}$$

The above mentioned mapping is reliable, robust with sufficient accuracy for estimating the focal length corresponding to a given zoom value. In Fig. 4(a), a graphical representation of this mapping is shown for employed camera.

Finally, we check the precision of the focal length estimation process for 12 zoom ticks which are not used in the estimation of zoom to focal length mapping. For each zoom tick, a pair of images are captured by tilting the PTZ camera and calibrated focal length $\bar{f}$ are found using image pair. Also, we interpolate the focal length f using (22) for all these selected zoom ticks. The absolute difference between these two focal lengths $|\bar{f} - f|$ are noted for each zoom values. A graphical representation of these difference with respect to selected zoom ticks are shown in Fig. 4(b). We found the mean of these differences is 0.12 mm which is acceptable in proposed zoom compensation algorithm. Thus, the above method is reliable for estimating the focal length for a given zoom value.

3.3. Image rectification

In order to estimate the two homographies H and H', the pairs of corresponding points are obtained using the SIFT-descriptor approach (Lowe, 2004) in each pair of images. An outlier detection process is performed to remove the outliers from matching features. Assume that the SIFT matching returns a number N of corresponding points (inliers) in the two images, i.e., $(x_i, x_i'), i = 1, 2, \ldots\ldots, N$. The two rectification transformations can be computed by minimizing the following cost function:

$$E(H, H') = \sum_{i=1}^{N} (x_i'^T H'^T F_\infty H x_i)^2 \tag{23}$$

Since the first row of F_∞ is a null vector, (23) involves only the second and third rows of H and H', and thus the first row of both H and H' cannot be reliably recovered by the minimization process. 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

$$d(H, H') = \sum_{i=1}^{N} (|(H x_i) - (H' x_i')|) \tag{24}$$

In other words we want to minimize the vertical distance between corresponding points. The nonlinear objective function of the optimization problem together with the constraints can be defined as

$$F(H, H') = E(H, H') + \eta d(H, H') \tag{25}$$

The minimization of (25) is a nonlinear minimization problem. To perform this minimization we choose the Levenberg–Marquardt algorithm because of its effectiveness. Before applying this process, we need to derive the Jacobian matrix of (25), i.e., $\partial F/\partial \phi$, where the vector ϕ contains all unknown parameters of H, H' and η. A good initial estimate is important to guarantee the convergence to a correct solution. To overcome this we use pyramidal search, i.e., first subsampled images are rectified, where the difference between coordinates is relatively small. Then, the estimated homographies from the subsampled level are used to initialize the estimation for the finer levels.

4. Results and discussions

Here, we first evaluate the performance of proposed rectification process using synthetic data. The correctness of proposed algorithm is checked by calculating the error in 3-D position of reconstructed points from rectified image pairs. Fig. 5 shows the result for rectification of an image pair having points in a elliptical geometry. The dark black points are used as the pairs of matching points in the rectification process. The focal ratio is taken 0.70 for this experiment for two images. Rectification has been made with and without making the heterogeneity compensation. It can be seen from the results that the rectified images (points in elliptical shape) obtained from direct algorithm are more distorted when compared to rectified images obtained by proposed two step

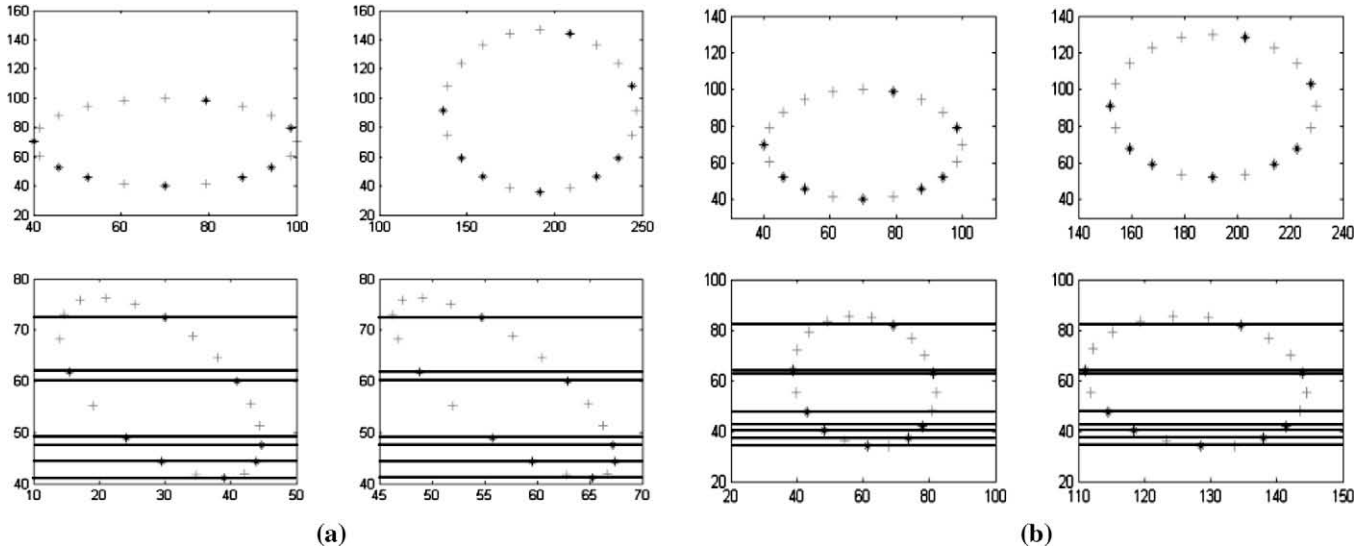

Fig. 5. Original (top) and rectified (bottom) pairs of stereo images: (a) direct rectification approach; (b) proposed rectification approach.

rectification algorithm (heterogeneity compensation and then rectification).

A quantitative comparison is made between the performances of proposed two-step rectification algorithm and direct rectification algorithm by considering different focal ratio for synthetic data. The 3-D position of reconstructed points are found from the both (proposed and direct) kinds of rectified pairs. A mean value and standard deviation value are calculated for the absolute difference in 3-D position between original and reconstructed points. Fig. 6 shows the results for the above described comparison. The horizontal axis represents the focal ratio between two images while the vertical axis represents the error. It is clearly visible from the given figure that the reconstruction error is significantly minor and more robust for lower focal ratios in the case of the proposed algorithm. The error increases drastically in the case of direct rectification algorithm with decreasing focal ratios.

To evaluate the performance of the proposed rectification algorithm in the case of real sequences, images from a parking lot are captured by an AXIS-213 PTZ network camera and an Axis-221 static network camera. The images captured by these cameras are of equal size 352×288. The zoom levels of these cameras have been acquired online and the focal length values are estimated from the zoom values using (22) after refining the zoom parameters. The aim of the proposed technique is to reliably rectify heterogeneous stereo images, i.e., when stereo images are acquired with different internal camera parameters. Therefore, different pairs of sequences having various focal ratios have been used for experimental study. SIFT are used to obtain pairs of corresponding points in the left and right images. The algorithm has been tested for a fixed number of matching points for all pairs of images.

We run a set of experiments by progressively reducing the focal ratio. In Fig. 7, the rectification results are shown for a pair of images having focal ratio 0.85 and rectified with 15 pairs of matching points. As it can be seen from these two pairs of images the performance improvement obtained by the proposed solution starts to be evident. The same comparison is carried out with Fig. 8, the rectification results have been shown for a pair of images having focal ratio 0.70 and rectified with 15 pairs of matching points. A quantitative comparison between the two approaches (proposed and direct) is given in Table 1. The error measure used in the comparison is the mean and standard deviation of the vertical distances between a fixed number of randomly chosen corresponding points:

$$r_i = |(Hx_i) - (H'x_i')|$$

Since rectification gives a pair of images in which corresponding epipolar lines should be collinear and parallel to horizontal axis, the above criterion is suitable for computing the error in a rectified pair of images in case of real data.

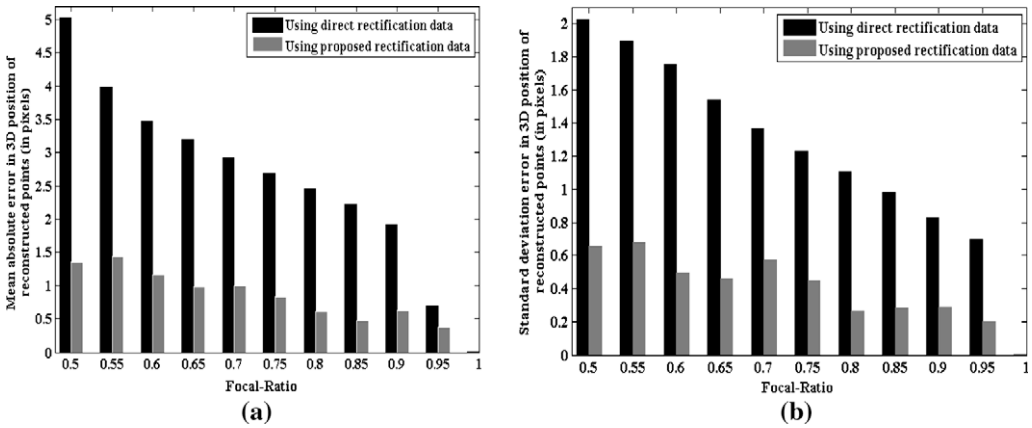

Fig. 6. Mean and standard deviation error in 3-D position of reconstructed points.

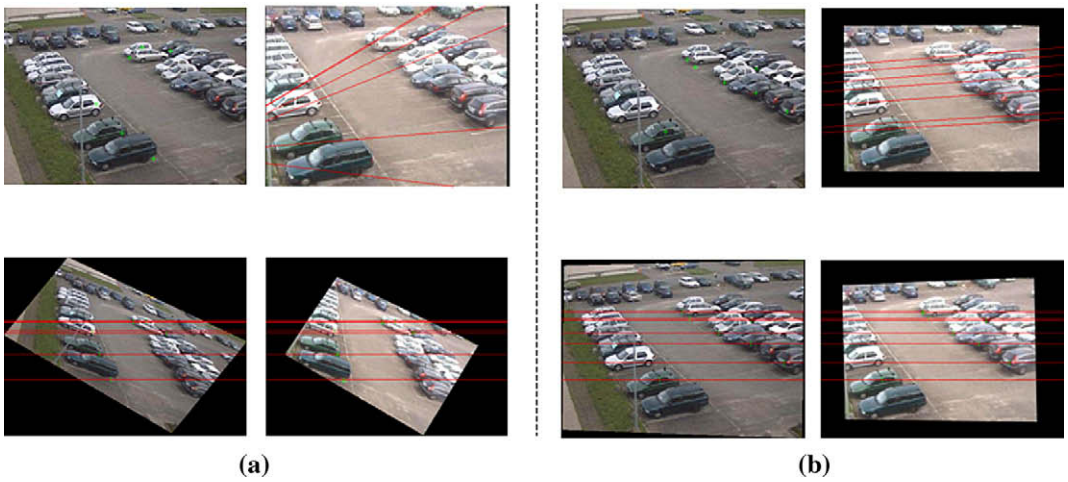

Fig. 7. (a) Direct, and (b) proposed rectification of a pair of stereo images having focal ratio 0.85. First row: pair of stereo images, second row: rectified pair of stereo images.

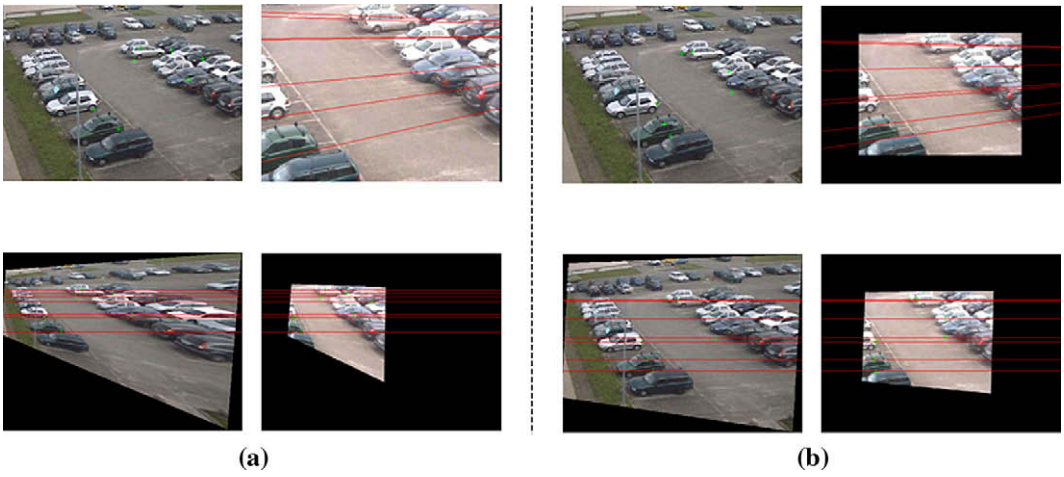

Fig. 8. (a) Direct, and (b) proposed rectification of a pair of stereo images having focal ratio 0.70. First row: pair of stereo images, second row: rectified pair of stereo images.

Table 1
Error (in pixels) in direct and proposed rectification algorithms corresponding to different focal ratios.

Focal ratio	Error in direct rectification		Error in proposed rectification	
	Mean	Standard deviation	Mean	Standard deviation
1.0	0.2305	0.1810	0.2305	0.1810
0.95	0.8723	0.5061	0.2706	0.2103
0.90	1.0543	0.5903	0.2835	0.2488
0.85	1.2537	1.1620	0.4651	0.3120
0.80	1.6956	1.2735	0.3967	0.3298
0.75	2.2137	1.9803	0.4505	0.3035
0.70	2.8253	1.5109	0.4003	0.2844
0.65	3.4200	2.5046	0.6341	0.4362
0.60	4.2723	3.8115	0.6885	0.3780
0.55	5.0572	4.1738	0.7617	0.4426
0.50	6.1910	5.9285	0.7745	0.5685

Table 1 shows that the error is very high when the rectification process has been performed directly without compensating the heterogeneity effect while the error is very small when using the proposed algorithm up to a focal ratio 0.50 (static camera zoom level 1× and PTZ camera zoom level 17×). When the focal ratio is reduced less than 0.50 then a gradual increment is noticed in the rectification error for both the approaches, since SIFT matching is not much reliable when the focal ratio is below 0.50, in complex scene like parking lots.

5. Conclusions

We have proposed a heterogeneity compensation based rectification algorithm for rectifying uncalibrated and non-homogeneous pairs of stereo images. An image shrinking and zero padding based heterogeneous compensation method has been employed to obtain pair of homogeneous images, both in terms of focal lengths and sizes. In order to achieve this result, the focal length of the PTZ camera has been interpolated by a mapping between focal length and refined zoom value. Finally, rectification transformations have been calculated for a homogeneous pair of images by minimizing the epipolar constraint using pairs of matching points found using SIFT descriptor. A constraint has been added in the objective function of nonlinear minimization problem to improve the accuracy of the proposed algorithm. The proposed algorithm has been tested for rectifying synthetic and real pairs of images. It has been found from the experimental study that the proposed algorithm performs better than a direct rectification approach.

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)". Sanjeev 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. Electron. Lett. 36 (5), 419–420.

Ayache, N., Hansen, C., 1988. Rectification of images for binocular and trinocular stereovision. In: IEEE Internat. Conf. on Pattern Recognition, Rome, Italy.

Ayache, N., Lustman, F., 1991. Trinocular stereo vision for robotics. IEEE Trans. Pattern Anal. Machine Intell. 13, 73–85.

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

Fusiello, A., Israra, L., 2008. Quasi epipolar unclaibrated rectification. In: IEEE Internat. Conf. on Image Process. (ICPR), pp. 1–4.

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

Gluckman, J., Nayar, S., 2001. Rectifying transformations that minimize resampling effects. In: IEEE Internat. Conf. on Comput. Vision and Pattern Recognition, Kauai, HI, USA.

Hartley, R., 1999. Theory and practice of projective rectification. Internat. J. Comput. Vision 35, 115–127.

Isgro, F., Trucco, E., 1999. On robust rectification for uncalibrated images. In: IEEE Internat. Conf. on Image Analysis and Process.

Loop, C., Zhang, Z., 1999. Computing rectifying homographies for stereo vision. In: IEEE Internat. Conf. on Comput. Vision and Pattern Recognition, Fort Collins, CO, USA.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Internat. J. Comput. Vision 60 (2), 91–110.

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

Pollefeys, M., Kock, R., Gool, L., 2001. A simple and efficient rectification method for general motion. In: IEEE Internat. Conf. on Image Process., Corfu, Greece.

Robert, L., Zeller, C., Faugeras, O., 1997. Applications of non-metric vision to some visually-guided robotics tasks. 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. In: Asian Conf. on Comput. Vision.

Wan, D., Zhou, J., 2008. Stereo vision using two PTZ cameras. Computer Vision and Image Understanding 112 (2), 184–194.

Wan, D., Zhou, J., 2009. Multiresolution and wide-scope depth estimation using a dual-PTZ-camera system. IEEE Trans. Image Process. 18 (3), 677–682.

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

Zahrani, A., Ipson, S., Haigh, J., 2004. Application of a direct algorithm for the rectification of uncalibrated images. Inform. Sci. 160, 53–71.

Zhang, Z., 1998. Determining the epipolar geometry and its uncertainty: A review. Internat. J. Comput. Vision 27, 161–195.