Affine Transform Based Image Rectification For Better Disparity From Stereo Matching

Sunil Kumar Kopparapu
Cognitive System Research Laboratory, Tata Consultancy Services Limited, Navi Mumbai,
SunilKumar.Kopparapu@TCS.Com

Abstract*— Most stereo algorithms assume images to be epipolar aligned. There are two ways of achieving this (i) physically aligning the cameras or (ii) rectifying the stereo images after capturing them. For real-time or fixed stereo head applications aligning the cameras is preferable because rectifying the stereo images would require precious computational resources and once set the cameras remain in alignment. For applications involving mobility of stereo head, like in robots, it is preferable to rectify images using software because the alignment of cameras could change with time. In this paper we use affine like transform to rectify stereo images and demonstrate its usefulness in producing better disparity estimates and show it can be used to capture stereo images from a single camera.Most stereo algorithms assume images to be epipolar aligned. There are two ways of achieving this (i) physically aligning the cameras or (ii) rectifying the stereo images after capturing them. For real-time or fixed stereo head applications aligning the cameras is preferable because rectifying the stereo images would require precious computational resources and once set the cameras remain in alignment. For applications involving mobility of stereo head, like in robots, it is preferable to rectify images using software because the alignment of cameras could change with time. In this paper we use affine like transform to rectify stereo images and demonstrate its usefulness in producing better disparity estimates and show it can be used to capture stereo images from a single camera.

I. Introduction

Estimation of disparity and hence depth from stereo images is an important prerequisite in many computer vision and robotics application. Disparity is defined as the *spatial differences* between the corresponding points in the scene, captured by the left and the right camera. Let $L(x, y)$ and $R(x, y)$ be the images captured by the left and the right camera respectively (x, y represent the spatial coordinates of the image). If $L(x_l, y_l)$ in the left image and $R(x_r, y_r)$ in the right image correspond to the same object in the scene, then the disparity is estimated as $\mathcal{D}(x, y) = (\mathcal{D}(x), \mathcal{D}(y))$, where $\mathcal{D}(x) = x_l - x_r$ and $\mathcal{D}(y) = y_l - y_r$ There are several approaches suggested to estimate disparity ([1], [2], [3], [4], [5], [6], [7], [8], [9] to cite a few) and most of the stereo algorithms assume that the stereo image pair are epipolar aligned, namely, $y_l = y_r$ or $\mathcal{D}(y) = 0$, so that *spatial differences* between the images in only in one direction and hence the disparity estimation problem reduces to a 1-D search problem. While this assumption on one hand significantly reduces the computational complexity while estimating the disparity but on the other hand puts an excessive restriction on alignment of stereo camera pair. There are essentially two ways of achieving epipolar alignment in the stereo images: (i) physically aligning the cameras [10] or (ii) rectifying the stereo images [11] after capturing them (in software). Which of these to use depends on the application in hand. For example in real-time applications where time is a major constraint, it is best to physically align the cameras to avoid rectifying the images to epipolar align them before estimating disparity. On the other hand in applications where the stereo camera pair is constantly subjected to erratic movement (example a robot in a rough terrain) it is preferable to align the images after they have been acquired because the alignment of the cameras mounted on the stereo head could change because of jerky movement.

In this paper, we assume an application of the type where the mounted stereo head is constantly in motion. In such situations there is a good chance of the alignment of the cameras on the stereo head to change with time and hence look at rectifying stereo images after acquiring them. In Section II we first show the need for epipolar geometry correction because of camera misalignment and then present a scheme for rectifying stereo images based on *affine like* transform. In Section III show the effect of such a rectification on estimating disparity from stereo image pair. We further show that the rectification procedure enables produce better disparity estimates from stereo image pairs on a stereo pair of images captured from a single off the shelf digital camera. The stereo image pair using the single digital camera is generated by first capturing a scene and a a second snap of the same scene is captured by physically moving the camera a little to the right.

II. Image Rectification using Affine like Transform

Image rectification can be loosely defined as a process of rectifying the image to take care of geometrical distortion creeping into the imaging system. Image rectification is used in a big way in remote sensing where geometrical distortion are significant because of the use of airborne vehicles used to capture the scene. In general, geometrical distortion can be taken care of provided we know the transformation that caused the distortion. It would be a case of finding the inverse transform (assuming it exists) of the transformation that caused the geometrical distortion.

To appreciate the need for image rectification stereo image pair, observe that for an off the shelf CCD camera with a lens of 8 mm focal length and pixel size of 11 μm, an error of approximately 1^o about the optical axis results in a shift of 0.16 mm ($\tan(1^o) \times 8$). This translates to approximately

14 $(0.16 \times 10^{-3}/11 \times 10^{-6})$ pixels across the line which violates the epipolar line constraint and is for most stereo vision applications unacceptable. Image rectification in stereo images would essentially mean epipolar alignment of the images namely the point corresponding to the same scene point should lie on the same line in both the left (L) and the right (R) images. The rectification can be handled provided we know the relative orientation of the two cameras producing the stereo image pair. In practice this would involve knowing the relative orientation of the cameras for each set of stereo pair. This is a task of *camera calibration* [12] [13] and is a laborious process especially if the relative orientation of the cameras change frequently with time.

In this paper, we assume that (a) we have no *apriori* knowledge of the relative orientation of the cameras on the stereo head that produced the stereo image pair and (b) that we are able to identify distinct points corresponding to the same scene point in the two stereo images *fairly* accurately.

Denote N distinct points in the left image by $\{(x_l^1, y_l^1), (x_l^2, y_l^2), \cdots (x_l^N, y_l^N)\}$ and the corresponding N points in the right image by $\{(x_r^1, y_r^1), (x_r^2, y_r^2), \cdots (x_r^N, y_r^N)\}$. The corresponding points in the left and the right image represent the same point in the 3-D scene. We can geometrically rectify the images so that the same point in the scene have the same y-ordinate in the left and the right images. Consider the following transformation

$$y_l^i = a_0 + a_1 x_r^i + a_2 y_r^i \quad \text{for} \quad i = 1, 2, \cdots N$$

This can be written in the matrix form as

$$L = RA \tag{1}$$

where,

$$L = \left[y_l^1 \ y_l^2 \ \cdots \ y_l^N\right]^T, \quad A = [a_0 \ a_1 \ a_2]^T$$

and

$$R = \begin{bmatrix} 1 & y_r^1 & x_r^1 \\ 1 & y_r^2 & x_r^2 \\ \vdots & \vdots & \vdots \\ 1 & y_r^N & x_r^N \end{bmatrix}$$

The transformation coefficients can be obtained by solving the matrix equation (1). Lets assume that $N > 3$, namely, we have atleast 3 corresponding points in both the left and the right stereo image pair. This results in solving the matrix equation where there are more number of equations than the number of variables (a_0, a_1, a_2). For this case we need to settle for a minimum least squares norm solution, namely

$$\min_A \|L - RA\|_2^2$$

The least square norm solution is given by [14]

$$A = (R^T R)^{-1} R^T L \tag{2}$$

The transformation coefficients A can be used to find the coordinates of the rectified right image as shown in Listing 1. The rectified image coordinates $(\hat{x}_r, \hat{y}_r)$ is used to construct the rectified right image $\hat{R}$ (see Listing 2). Now, L and $\hat{R}$ are epipolar aligned and can be used to estimate disparity. In all our experiments we have used a method based on sum of absolute differences (SAD) to determine disparity[1] from the stereo pair [15].

Listing 1 Finding coordinates of the rectified right image

for i=1 to rows **do**
  for j=1 to cols **do**
    $\hat{x}_r(i,j) = x_r(i,j)$
    $\hat{y}_r(i,j) = a_0 + a_1 y_r(i,j) + a_2 x_r(i,j)$
    $\{a_0, a_1, a_2$ are obtained from (2).$\}$
  end for
end for

Listing 2 Rectified right image

for i=1 to rows **do**
  for j=1 to cols **do**
    $\hat{R}(i,j) = R(\hat{x}_r(i,j), \hat{y}_r(i,j))$
    $\{R$ is the right image and the $\hat{R}$ is the rectified image $\}$
    $\{\hat{x}_r(i,j), \hat{y}_r(i,j)$ are obtained from Listing 1 $\}$
  end for
end for

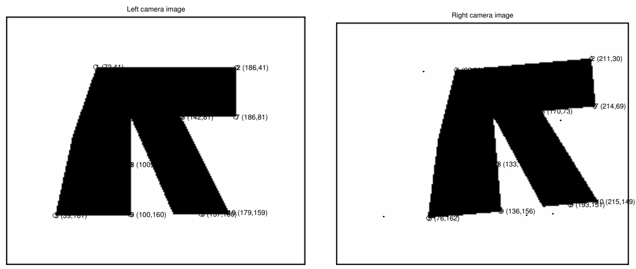
III. Experimental Results

Initially we show that the proposed *affine like* transform (Section II) rectifies the stereo image pair to epipolar align them and later we show that the rectification helps in producing better disparity map for real images captured using a single camera.

Experiments were carried out on synthetically generated data (Fig. 1) to check the correctness of rectification process. The right image was obtained by shifting the pixels in the left image (Fig. 1(a)) along the x axis by a constant amount (this amounts to the assumption that the image in the scene is at a constant distance from the stereo camera pair). In addition, the right image (Fig. 1(b)) is subjected to a small rotation; assumed due to camera misalignment. The rectification process described in the paper should take care of the rotation so that the images are epipolar aligned.

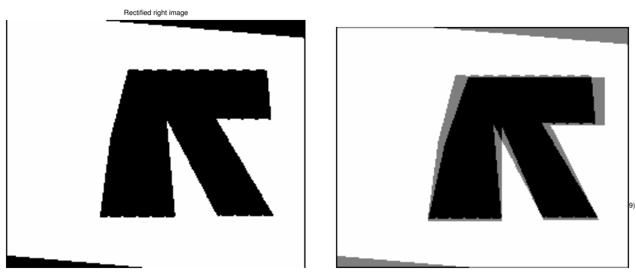
For purpose of demonstration we choose 10 points in the left and corresponding 10 points in the right image. The chosen points are marked in Fig. 1. The procedure described in Section II results in a rectified image as shown in Fig. 2(a). It can be seen that the right image has been rectified such that the rotation has been taken care. Fig. 2(b) shows the superimposition of Fig. 1(a) and Fig. 2(a). Note that the rectified image does not loose its depth defining information (constant shift along the x-axis). The medium gray level is the region where the rectified image (Fig. 2(a)) and the actual right image (no rotation just constant shift of the left image) mismatch. This shows that the image has been modified to take care of the possible misalignment in the stereo camera pair.

[1] In this paper due to paucity of space we will not discuss the procedure adopted to extract disparity information from stereo image pair



(a) Left camera image. (b) Right camera image.

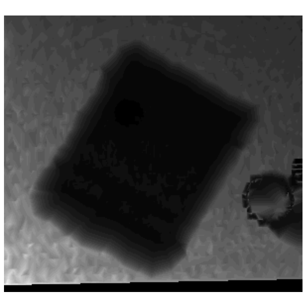
Fig. 1. A synthetic stereo image pair.



(a) Rectified using *affine like* transform. (b) Rectified right image (Fig. 1(b)).

Fig. 2. Stereo image pair rectification.

(a) Left image. (b) Right image.



(c) Disparity to generate Fig. 3(b).

Fig. 3. Stereo pair using known disparity.

(a) Unrectified image. (b) Disparity map.

Fig. 4. Disparity without rectification.

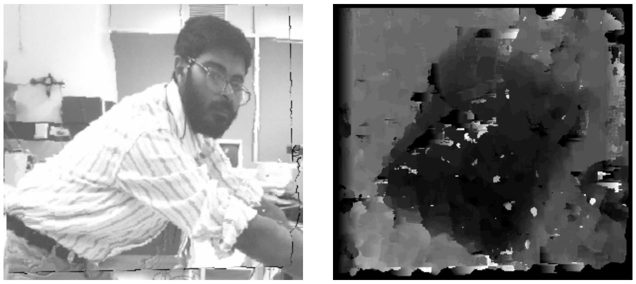
In the next set of results we show how the process of stereo image rectification can improve disparity estimation. Fig. 3 shows the generation of the stereo image pair. Fig. 3(a) is a 8 bit gray image captured using Sony CCD camera and Fig. 3(c) is the disparity map which was used to create the right image of the stereo pair (Fig. 3(b)). The right image was rotated by 2^o in the counter clock wise direction (Fig. 4(a)) to simulate distortion due to camera misalignment. Fig. 4(b) shows the disparity estimated using sum of absolute differences (SAD) metric [16] on the left image (Fig. 3(a)) and the unrectified right image (Fig. 4(a)). The rectification scheme detailed in Section II is used to rectify Fig. 4(a) and the resulting rectified image is displayed in Fig. 5(a). The left image (Fig. 3(a)) and the rectified right image are used to calculate the disparity which is shown in Fig. 5(b). Clearly, the disparity map after the process of rectification is much better than the disparity map before image rectification (compare Fig. 4(b) and Fig. 5(b)).

A. Application: Hand held digital camera to capture stereo images

The next set of results is useful to build an affordable stereo camera. We show the use of the rectification scheme to extract depth from images captured from a single digital camera. A hand held stereo camera head can be best described as a single off the shelf digital camera used to gather stereo pair of images. A scene of interest is first captured by a person and then a second snap of the same scene is captured by physically moving the camera a little to the right. This results in a pair of stereo images. Clearly, the stereo pair are not epipolar aligned.

Fig. 6(a) was captured by taking the image of a scene and then physically shifting the camera (in 3-D space) to the right by a small amount to gather a second image (Fig. 6(b)). Fig. 6(a) and 6(b) correspond to the left and the right images of the stereo pair taken by a *hand held* digital camera[2] (tripod was not used). The scheme suggested in this paper was used to rectify the right image which is shown in Fig. 7. Fig. 8(a) and 8(b) show the disparity map using the SAD measure. Fig. 8(a) is the disparity map estimated from the original left and

[2]Sony Digital camera Maravik was used



(a) Rectified image. (b) Disparity map.

Fig. 5. Disparity after rectification.

(a) Left Image. (b) Right Image.

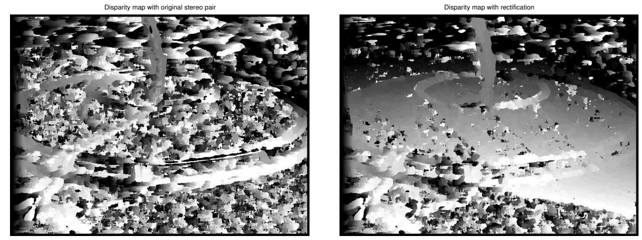
Fig. 6. Stereo using a hand held camera.

the right image (unrectified) and the disparity map in Fig. 8(b) is estimated using the left and the rectified right image. It can be observed that rectification using the scheme suggested in this paper significantly improves the disparity map (compare Fig. 8(b) to 8(a)).

IV. Conclusions

This paper describes a scheme based on *affine like* transform to rectify stereo image pair to produce epipolar aligned images. The scheme can be used as a preprocessor to stereo algorithms which need the images to be epipolar aligned. We have demonstrated how the suggested scheme can help in getting a significantly improved disparity map even when one uses a hand held camera to generate stereo images (affordable stereo camera).

Fig. 7. Rectified Right Image.



(a) Disparity (unrectified). (b) Disparity (rectified).

Fig. 8. Disparity estimation.

V. References

[1] S. T. Barnard and W. B. Thompson, "Disparity analysis of images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 2, pp. 330 – 340, 1980.
[2] D. Marr and T. Poggio, "Cooperative computation of stereo disparity," *Science*, vol. 194, pp. 283–287, 1976.
[3] ——, "A computational theory of human stereo vision," in *Procedings Royal Society London*, 1979.
[4] S. Pollard, J. Mayhew, and J. Frisby, "PMF: A stereo correspondence algorithm using a disparity gradient limit," *Perception*, vol. 14, pp. 449 – 470, 1985.
[5] W. E. L. Grimson, "Computational experiments with feature based stereo algorithm," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 7, pp. 17 – 34, 1985.
[6] K. Prazdny, "Detection of binocular disparities," *Biological Cybernatics*, vol. 52, pp. 93 – 99, 1985.
[7] R. D. Eastman and A. M. Waxman, "Using disparity functionals for stereo correspondence and surface reconstruction," *CVGIP: Image Understanding*, vol. 39, pp. 73–101, 1987.
[8] N. M. Nasarabadi and C. Y. Choo, "Hopfield network for stereo vision correspondence," *IEEE Tran. Neural Networks*, vol. 1, pp. 5 – 13, 1992.
[9] K. S. Kumar and U. B. Desai, "New algorithms for 3D depth estimation from binocular stereo," *Journal of the Franklin Institute*, vol. 331B, pp. 531 – 554, 1994.
[10] W. Zhao and N. Nandhakumar, "Effects of camera alignment errors on stereoscopic depth estimates," *PR*, vol. 29, pp. 2115–2126, 1996.
[11] D. V. Papadimitriou and T. J. Dennis, "Epipolar line estimation and rectification for stereo images pairs," *IEEE Transactions on Image Processing*, vol. 3, no. 4, pp. 672–676, April 1996.
[12] R. Tsai, "A versatile camera calibration technique for high accuracy 3-D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE Journal of Robotics and Automation*, vol. 3, no. 4, pp. 323–344, Aug. 1987.
[13] S.-W. Shih, P. Hung, and W.-S. Lin, "When should we consider lens distortion in camera calibration," *Pattern Recognition*, vol. 28, no. 3, pp. 447–461, 1995.
[14] G. H. Golub and C. F. V. Loan, *Matrix Computations*. The Johns Hopkins University Press, 1996.
[15] J. Banks, M. Bennamoun, K. Kubik, and P. I. Corke, "Evaluation of new and existing confidence measures for stereo matching," in *IVCNZ*, 1998, pp. 252–261.
[16] H. Kim, D. B. Min, S. Choi, and K. Sohn, "Real-time disparity estimation using foreground segmentation for stereo sequences," *Optical Engineering*, vol. 45, no. 3, March 2006.