Three-step image rectification

Pascal MONASSE[1]
monasse@imagine.enpc.fr
Jean-Michel MOREL[2]
morel@cmla.ens-cachan.fr
Zhongwei TANG[2]
tang@cmla.ens-cachan.fr

[1] IMAGINE, LIGM/Université Paris-Est/École des Ponts ParisTech
Marne-la-Vallée, France
[2] CMLA, ENS-Cachan
Cachan, France

Image stereo-rectification is the process by which two images of the same solid scene undergo homographic transforms, so that their corresponding epipolar lines coincide and become parallel to the x-axis of image. A pair of stereo-rectified images is helpful for dense stereo matching algorithms. It restricts the search domain for each match to a line parallel to the x-axis. Due to the redundant degrees of freedom, the solution to stereo-rectification is not unique and actually can lead to undesirable distortions or be stuck in a local minimum of the distortion function.

In this paper a robust geometric stereo-rectification method by a three-step camera rotation is proposed and mathematically explained. Unlike other methods [1, 2, 3, 4, 5] which reduce the distortion by explicitly minimizing an empirical measure, the intuitive geometric camera rotation angle is minimized at each step. For un-calibrated cameras, this method uses an efficient minimization algorithm by optimizing only one natural parameter, the focal length. This is in contrast with all former methods which optimize between 3 and 6 parameters.

The fundamental matrix corresponds to two stereo-rectified images if and only if it has the special form (up to a scale factor)

$$[\mathbf{i}]_\times = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}_\times = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}. \qquad (1)$$

And the rectification can be achieved by applying a homography $\mathbf{H}$ on the image. This homography can be written in the form: $\mathbf{H} = \hat{\mathbf{K}}\mathbf{R}\mathbf{K}^{-1}$, where $\mathbf{R}$ is the relative rotation before and after rectification, and $\mathbf{K}$, $\hat{\mathbf{K}}$ calibration matrix before and after rectification.

Given a group of non-degenerate correspondences between two images, the fundamental matrix $\mathbf{F}$ can be computed and also two corresponding epipoles $\mathbf{e} = (e_x, e_y, 1)^T$ ($\mathbf{Fe} = 0$), $\mathbf{e}' = (e'_x, e'_y, 1)^T$ ($\mathbf{e}'^T\mathbf{F} = 0$). Assume cameras are weakly calibrated and only unknown is focal length f. The idea is to transform both images so that the fundamental matrix gets the form $[\mathbf{i}]_\times$. Unlike the other methods which directly parameterize the homographies from the constraints $\mathbf{He} = \mathbf{i}$, $\mathbf{H}'\mathbf{e}' = \mathbf{i}$ and $\mathbf{H}'^T[\mathbf{i}]_\times\mathbf{H} = \mathbf{F}$ and find an optimal pair by minimizing a measure of distortion, we shall compute the homography by explicitly rotating each camera around its optical center. The algorithm is decomposed into three steps (Fig. 1):

1. Compute homographies $\mathbf{H}_1$ and $\mathbf{H}'_1$ by rotating both cameras respectively so that the left epipole $(e_x, e_y, 1)^T$ is transformed to $(e_x, e_y, 0)^T$ and the right epipole $(e'_x, e'_y, 1)^T$ to $(e'_x, e'_y, 0)^T$.

2. Rotate both cameras so that $(e_x, e_y, 0)^T$ is transformed to $(1, 0, 0)^T$ and $(e'_x, e'_y, 0)^T$ to $(1, 0, 0)^T$. The corresponding homographies are denoted by $\mathbf{H}_2$ and $\mathbf{H}'_2$.

3. Rotate one camera or both cameras together to compensate the residual relative rotation between both cameras around the baseline. The corresponding homographies are denoted by $\mathbf{H}_3$ and $\mathbf{H}'_3$.

Given a focal length f, in the first two steps, the camera rotation matrix is computed by minimizing the rotation angle, which reduces distortion. At the third step, the residual fundamental matrix is modified such that it is compatible with the given calibration matrix. Thus the residual rotation matrix around the baseline between two cameras can be extracted by SVD. This three-step decomposition gives a parametrization of the modified fundamental matrix $\tilde{\mathbf{F}}$ in f. The best f is found by minimizing the distances from the points to the corresponding epipolar lines: $S(f) = \sum_{i=1}^{N} d(\mathbf{x}'_i, \tilde{\mathbf{F}}\mathbf{x}_i) + d(\mathbf{x}_i, \tilde{\mathbf{F}}^T\mathbf{x}'_i)$.

This three-step algorithm has the advantage of reducing the distortion implicitly. Comparative experiments show that the algorithm has an accuracy comparable to the state-of-art, but finds the right minimum in cases where other methods fail, namely when the epipolar lines are far from horizontal(Fig. (2).

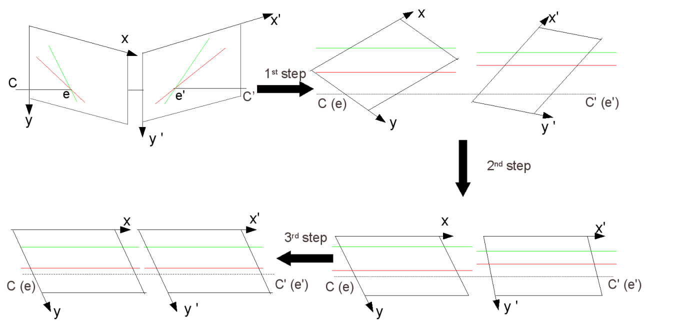

Figure 1: Three-step rectification. First step: the image planes become parallel to CC'. Second step: the images rotate in their own plane to have their epipolar lines also parallel to CC'. Third step: a rotation of one of the image planes around CC' aligns corresponding epipolar lines in both images. Note how the pairs of epipolar lines become aligned.

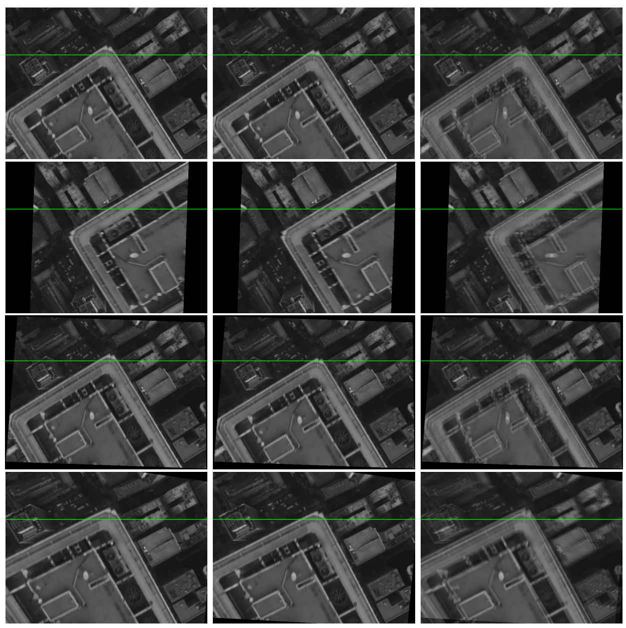

Figure 2: Image pair "Building" rectified by different methods. From top to bottom: original images, proposed method, Hartley method [3] and Fusiello *et al.* method [1]. A horizontal line is added to images to check the rectification. The third column represents an image average of each pair.

[1] A. Fusiello and L. Irsara. Quasi-Euclidean uncalibrated epipolar rectification. *ICPR*, pages 1–4, 2008.

[2] J. Gluckman and S.K. Nayar. Rectifying transformations that minimize resampling effects. *CVPR*, 1, 2001.

[3] R.I. Hartley. Theory and practice of projective rectification. *International Journal of Computer Vision*, 35(2):115–127, 1999.

[4] C. Loop and Z. Zhang. Computing rectifying homographies for stereo vision. *Computer Vision and Pattern Recognition*, 1:125–131, 1999.

[5] J. Mallon and Paul F. Whelan. Projective rectification from the fundamental matrix. *Image and Vision Computing*, 23:643–650, 2005.