

Quasi-Euclidean Epipolar Rectification of Uncalibrated Images

Andrea Fusiello · Luca Irsara

Received: date / Accepted: date

Abstract This paper deals with the problem of epipolar rectification in the uncalibrated case. First the calibrated (Euclidean) case is recognized as the ideal one, then we observe that in that case images are transformed with a collineation induced by the plane at infinity, which has a special structure. Hence, that structure is imposed to the sought transformation while minimizing a rectification error. Experiments show that this method yields images that are close to the ones produced by Euclidean rectification.

1 Introduction

Epipolar rectification is an important stage in dense stereo matching, as almost any stereo algorithm requires rectified images, i.e., images where epipolar lines are parallel and horizontal and corresponding points have the same vertical coordinates. Whereas in 3D reconstruction the cameras are usually calibrated, in other applications such as view-synthesis [12] and parallax-based mosaics [10] the internal parameters are unknown.

In the case of calibrated cameras the *Euclidean* epipolar rectification is unique up to trivial transformations [2]. On the contrary, in the case of uncalibrated cameras, there are more degrees of freedom in choosing the rectifying transformation [7] and a few competing methods are present in the literature [11, 13]. Each aims at producing a “good” rectification by minimizing a measure of distortion, but none is clearly superior to the others, not to mention the fact that

there is no agreement on what the distortion criterion should be.

Above all, none of them achieves results comparable to the Euclidean epipolar rectification, which can be indisputably taken as the target result. This paper aims at achieving a good approximation of the Euclidean epipolar rectification, which we refer to as *quasi-Euclidean* epipolar rectification.

Geometrically, in the Euclidean frame, rectification is achieved by a suitable rotation of both image planes. The corresponding image transformation is the collineation induced by the plane at infinity. As a result, the plane at infinity is the locus of zero-disparity in the rectified stereo pair. This is signified by saying that Euclidean rectification is done with respect to the plane at infinity. Rectification is always referred to a plane, but in the uncalibrated case the reference plane is generic, for any plane can play the role of the infinity plane in the projective space. Our quasi-Euclidean rectification can be seen as referred to a plane that approximates the plane at infinity.

1.1 Previous work

The first work on uncalibrated rectification (called “matched-epipolar projection”) is [4], followed by [7], where the author tidies up the theory. He uses the condition that one of the two collineations should be close to a rigid transformation in the neighborhood of a selected point, while the remaining degrees of freedom are fixed by minimizing the distance between corresponding points (disparity). Along the same line, [17] also proposes a distortion criterion based on simple geometric heuristics.

The state-of-the-art papers in uncalibrated rectification [13, 11] dates back to 1999. Loop and Zhang [13] decompose each collineation into similarity, shearing and projective factors and attempt to make the projective component

A. Fusiello
Dipartimento di Informatica, Università di Verona
Strada Le Grazie, 15 - 37134 Verona, IT
Tel.: +39-045-8027088 - Fax: +39-045-8027068
E-mail: andrea.fusiello@univr.it

L. Irsara
EURAC research, Viale Druso 1, 39100 Bolzano, IT
E-mail: luca.irsara@eurac.edu

“as affine as possible”. Isgrò and Trucco [11] build upon [7] and propose a method that avoids computation of the fundamental matrix, using the same distortion criterion as in [7]. The practice has shown that the rectification produced by these methods is not always satisfactory, if compared to results obtained in the calibrated case. Wu and Yu [9] argue that minimizing the disparity might be the cause of the problem, and propose a technique which is similar to [11] but uses a different distortion criterion derived from [13]. Rectification ends up in a non-linear minimization with six degrees of freedom. In [3] the transformation that best preserve the sampling of the original images is selected, by penalizing minification and magnification effects. A similar approach, described in [16], consist in favoring transformations close to orthogonal, which translates in evaluating the singular values of the Jacobian.

A different approach is followed in [1]: they design the collineations so as to minimize the relative distortion between the rectified images (instead of the distortion of each rectified image with respect to the original one), and the remaining degrees of freedom are fixed by choosing the reference plane in the scene, that will have zero disparity in the rectified pair. This choice, which affects sensibly the quality of rectification, is left to the user. The merit of this approach, however, is to make the role of the reference plane explicit.

2 Background

In this section we shall briefly recapitulate the theory of calibrated (or Euclidean) epipolar rectification; the reader is referred to [2] for more details.

Given two camera matrices P_{or} and P_{ol} , the idea behind rectification is to define two new *virtual* cameras P_{nr} and P_{nl} obtained by rotating the actual ones around their optical centres until focal planes become coplanar. The rectification method describes how to compute the new cameras. Then (we concentrate on the right camera, but the same reasoning applies to the left one), the rectifying transformation that is to be applied to the image is given by the 3×3 matrix:

$$H_r = P_{nr1:3} P_{or1:3}^{-1} \quad (1)$$

where the subscript denotes a range of columns.

It is easy to see that H_r is the collineation induced by the plane at infinity between the old and the new cameras, hence it can be written as:

$$H_r = K_{nr} R_r K_{or}^{-1} \quad (2)$$

where K_{or} and K_{nr} are the intrinsic parameters of the old and new camera respectively, and R_r is the rotation that is applied to the old camera in order to rectify it.

The rectified images are as if they were taken by a pair of cameras related by a translation along the baseline. Hence, the zero-disparity plane is at infinity.

3 Method

We shall henceforth concentrate on the uncalibrated case. We assume that intrinsic parameters are unknown and that a number of corresponding points $\mathbf{m}_\ell^j \leftrightarrow \mathbf{m}_r^j$ are available. The method follows the same line as in [11]: it seek the collineations that make the original points satisfy the epipolar geometry of a rectified image pair.

The fundamental matrix of a rectified pair has a very specific form, namely it is the skew-symmetric matrix associated with the cross-product by the vector $\mathbf{u}_1 = (1, 0, 0)$:

$$[\mathbf{u}_1]_\times = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \quad (3)$$

Let H_r and H_ℓ be the unknown rectifying collineations. When they are applied to the corresponding tie-points $\mathbf{m}_\ell^j, \mathbf{m}_r^j$ respectively, the transformed points must satisfy the epipolar geometry of a rectified pair, namely:

$$(H_r \mathbf{m}_r^j)^\top [\mathbf{u}_1]_\times (H_\ell \mathbf{m}_\ell^j) = 0, \quad (4)$$

The left-hand side of (4) is an algebraic error, i.e., it has no geometrical meaning, so we used instead the Sampson error [15, 5], that is a first order approximation of the geometric error. The matrix $F = H_r^\top [\mathbf{u}_1]_\times H_\ell$ can be considered as the fundamental matrix between the original images, therefore, in our case, the squared Sampson error for the j -th correspondence is defined as:

$$E_j^2 = \frac{(\mathbf{m}_r^{j\top} F \mathbf{m}_\ell^j)^2}{(F \mathbf{m}_\ell^j)_1^2 + (F \mathbf{m}_\ell^j)_2^2 + (\mathbf{m}_r^{j\top} F)_1^2 + (\mathbf{m}_r^{j\top} F)_2^2} \quad (5)$$

where $(\cdot)_i$ is the i^{th} component of the normalized vector.

As this equation must hold for any j , one obtains a system of non-linear equations $\{E_j = 0\}$ in the unknown H_r and H_ℓ . A least-squares solution can be obtained with the Levenberg-Marquardt algorithm, but the way in which H_r and H_ℓ are parameterized is crucial, and characterizes our approach with respect to the previous ones. We force the rectifying collineations to have the same structure as in the calibrated (Euclidean) case, i.e., to be collineations induced by the plane at infinity, namely

$$H_r = K_{nr} R_r K_{or}^{-1} \quad H_\ell = K_{nl} R_\ell K_{ol}^{-1}. \quad (6)$$

The old intrinsic parameters (K_{ol}, K_{or}) and the rotation matrices (R_ℓ, R_r) are unknown, whereas the new intrinsic parameters (K_{nl}, K_{nr}) can be set arbitrarily, provided that vertical focal length and vertical coordinate of the principal point are the same. Indeed, it is easy to verify that the matrix $K_{nr}^\top [\mathbf{u}_1]_\times K_{nl}$ is equal (up to scale) to $[\mathbf{u}_1]_\times$, provided that the second and third row of K_{nr} and K_{nl} are the same. Hence it is not necessary to include the matrices K_{nr} and K_{nl} in the parameterization.

Each collineation depends in principle on five (intrinsic) plus three (rotation) unknown parameters. The rotation of one camera along its X -axis, however, can be eliminated. Consider the matrix

$$F = K_{or}^{-T} R_r^T [\mathbf{u}_1]_{\times} R_{\ell} K_{ol}^{-1}. \quad (7)$$

Let R'_r and R'_ℓ be the same matrices as R_r and R_ℓ after pre-multiplying with an arbitrary (but the same for both) rotation matrix about the X -axis. It is easy to verify that $R_r^T [\mathbf{u}_1]_{\times} R_\ell = R_r'^T [\mathbf{u}_1]_{\times} R_\ell'$. Geometrically, this coincide with rotating a rectified pair around the baseline, which do not alter the rectification, but, in a real camera, it affects the portion of the scene that is imaged. Accordingly, we set to zero the rotation around the X -axis of the left camera.

We further reduce the number of parameters by making an educated guess on the old intrinsic parameters: no skew, principal point in the centre of the image, aspect ratio equal to one. The only remaining unknowns are the focal lengths of both cameras. Assuming that they are identical¹ and equal to α , we get:

$$K_{or} = K_{ol} = \begin{bmatrix} \alpha & 0 & w/2 \\ 0 & \alpha & h/2 \\ 0 & 0 & 1 \end{bmatrix} \quad (8)$$

where w and h are width and height (in pixel) of the image.

In summary, the two collineations are parameterized by six unknowns: five angles and the focal length α . As suggested in [8], the focal length is expected to vary in the interval $[1/3(w+h), 3(w+h)]$, so we consider instead the variable $\alpha' = \log_3(\alpha/(w+h))$ which varies in $[-1, 1]$.

The minimization of the cost function (Program 1) is carried out using Levenberg-Marquardt, starting with all the unknown variables set to zero. When α' converges outside the boundaries of the interval $[-1, 1]$ a random restart is attempted. If the problem persist the minimization is carried out with fixed $\alpha' = 0$.

Finally, the new intrinsic parameters (K_{nr} and $K_{n\ell}$) are set equal to the old ones: $K_{nr} = K_{n\ell} = K_{ol}$, modulo a shift of the principal point, that might be necessary to centre the rectified images in the customary image coordinate frame. Horizontal translation has no effect on the rectification, whereas vertical translation must be the same for both images.

4 Results

In order to allow for comparison, we first tested our quasi-Euclidean rectification on the SYNTIM images, that have been widely used in the rectification literature. Then we evaluated the quasi-Euclidean rectification on the VSG images,

¹ The method could simply cope with two different focal length, at the cost of increasing by one the unknown parameters. However, we preferred to focus on the more common situation of having a stereo pair taken with two identical cameras.

Program 1 The quasi-Euclidean rectification cost function.

```
function err = costRectif(a,w,h,ml,mr) ;
%costRectif Compute rectification cost
%
% a is a vector of six elements containing the
% independent variables of the cost function,
% i.e., five rotation angles (Y-left, Z-left,
% X-right, Y-right, Z-right) and the focal length.
% w,h are the image width and height, respectively.
% ml,mr contain the corresponding image points.

yl=a(1); zl=a(2); xr=a(3);
yr=a(4); zr=a(5); f=3^a(6)*(w+h);

% estimate of the intrinsic parameters
% of the old cameras
Kol = [f, 0, w/2; 0, f, h/2; 0, 0, 1];
Kor = Kol;

% eulR applies rotations in the order Y-Z-X
Rl = eulR([0,yl,zl]); Rr = eulR([xr,yr,zr]);

% fundamental matrix btw original points
F = inv(Kor)'*Rr'*hat([1 0 0])*Rl*inv(Kol);

% compute Sampson error
err = sqrt(sampson(F,ml,mr));
```

that have been recently used in [16] and made publicly available by the authors. Eventually we tested our algorithm also on casual images taken by the authors, where a fully automatic procedure has been deployed.

SYNTIM images. Few (10-20) corresponding points had been manually selected in each image pair, and fed to the quasi-Euclidean rectification algorithm that automatically produced the rectified pairs shown in Fig. 1 and 2.

Calibration data (provided with the images) were used for computing the ground truth Euclidean rectification with the algorithm described in [2]. As the reader can appreciate in Fig. 1 and 2, our results are remarkably close to the Euclidean rectification. Unfortunately, a quantitative measure of distortion that captures the desired behaviour of rectification in every respect does not exist. Therefore, we are forced to resort to qualitative assessment.

In [11] the rectified images (reported in Fig:3) are in some cases very distorted and distinctly dissimilar from the Euclidean case (*Rubick*, *BalMire*, *BalMousse*). In the results shown in [9] the amount of distortion is qualitatively comparable to ours, but our results are still closer to the Euclidean rectification.

Table 1 reports the rectification error, computed as the average vertical distance between corresponding points:

$$e_{\text{rec}} = \frac{1}{N} \sum_j \sqrt{\left((H_r \mathbf{m}_r^j)_2 - (H_\ell \mathbf{m}_\ell^j)_2 \right)^2} \quad (9)$$



Fig. 1 Euclidean (left) and quasi-Euclidean rectification (right) of the SYNTIM pairs.

where N is the number of corresponding pairs. The figures, beside confirming that the pairs are indeed rectified, reveal that the error achieved by the quasi-Euclidean rectification are consistently smaller than in the Euclidean rectification. This makes sense, because the former explicitly minimizes a rectification error (although not exactly e_{rec}), whereas the latter derives the rectifying collineations directly from camera matrices in closed form. Therefore small calibration errors propagates to the rectification results without feedback.

This observation precludes the possibility of achieving uncalibrated rectification by performing autocalibration followed by Euclidean rectification, for autocalibration from only two images have been reported to be unstable [6, 18], hence parameters are far less accurate than those obtained from calibration. This implies that the rectification error would be macroscopic.

Table 1 Rectification errors for the SYNTIM pairs.

Images	quasi-Euc.	Euc.
	e_{rec}	e_{rec}
<i>Aout</i>	0.717	2.629
<i>BalMire</i>	0.932	0.985
<i>BalMouss</i>	0.547	6.237
<i>BalInria</i>	0.571	1.429
<i>Color</i>	0.583	0.782
<i>Rubik</i>	0.330	1.705
<i>Sport</i>	0.777	0.940
<i>Tot</i>	0.540	0.768

We noted experimentally that, considering in turn only the angle parameters and the focal length as the independent variables, the location of the minimum of the cost function is fairly insensitive to the value of the focal length (i.e., the

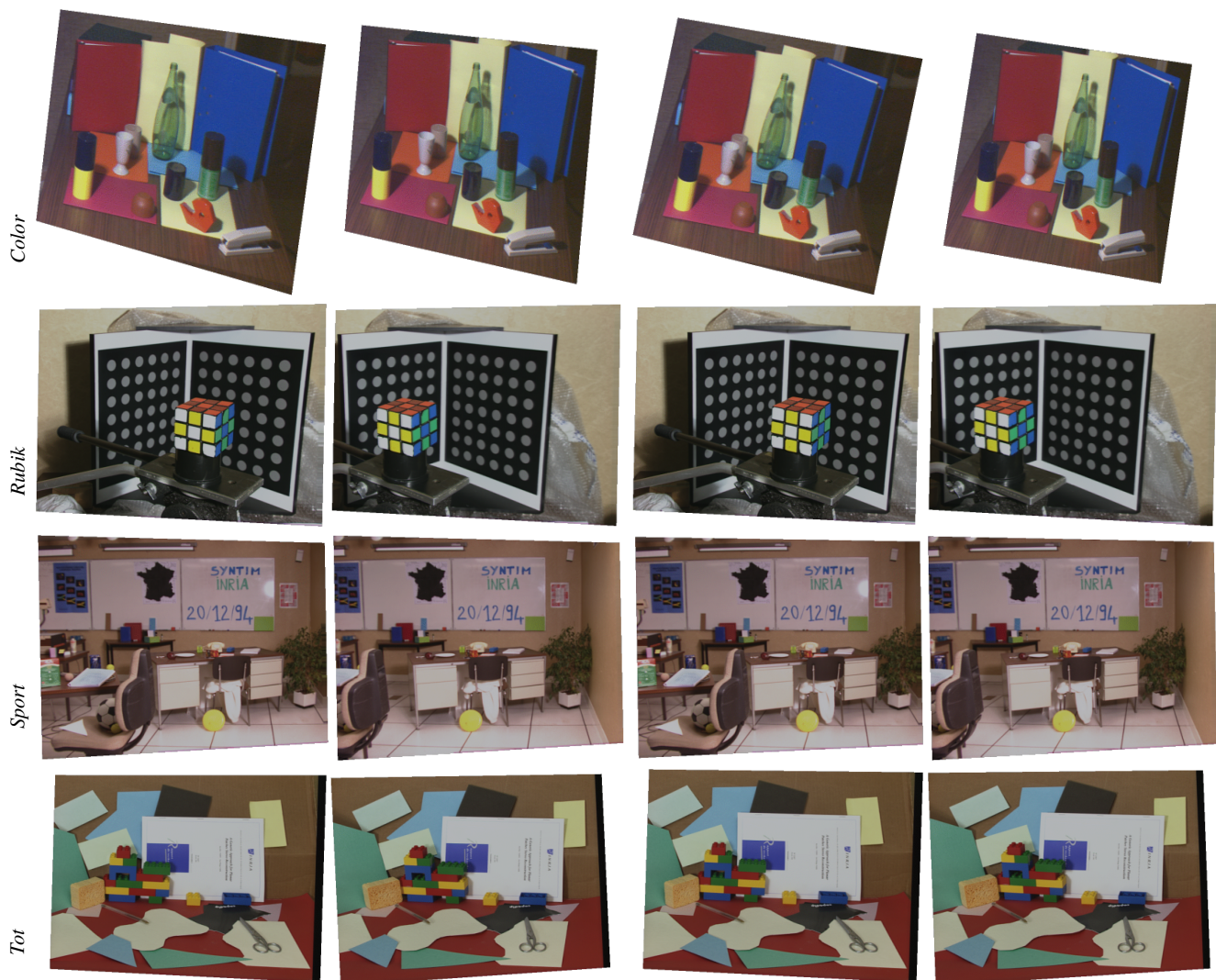


Fig. 2 SYNTIM pairs. Euclidean (left) and quasi-Euclidean (right) rectification.

correct angles can be obtained even when α is far away from the ground truth), whereas, in certain cases, small perturbations of the angles can drift the minimum away from the ground truth for the focal length (without affecting the quality of rectification, though). This seems to be related to the fact that computing the focal length from two uncalibrated views [6, 18] is ill conditioned when the cameras are verging, which is often the case in this stereoscopic pairs.

VSG images. In this experiments we compared our quasi-Euclidean rectification with the method proposed in [16], using the same metrics and the same images. Results are shown in Fig. 4. As can be noted by comparing the figures reported in Tab.2 with those reported in [16](Tab. 1) our algorithm consistently performs better than [16], which in turn have been shown to yield less distortion than [13] and [7].

Orthogonality and aspect ratio factors are defined in [16]: they attempt to measure the skewedness and relative scale

Table 2 Rectication error and distortion metrics (Orthogonality and Aspect ratio) for the VSG pairs. The values of the distortion metrics are the average of the left and right images.

Images	e_{rec}	Ortho.	Asp. rt.
<i>Lab</i>	0.104	90.077	1.003
<i>Slate</i>	0.119	90.019	1.001
<i>Roof</i>	0.669	90.093	1.002
<i>Arch</i>	0.211	90.116	1.004
<i>Yard</i>	0.359	90.046	1.002
<i>Drive</i>	0.487	90.037	1.002

change introduced by rectification. The ideal values are 90 and 1, respectively. The reader might appreciate that none of the metrics above have been explicitly optimized in our algorithm.

Casual images. A few other examples of quasi-Euclidean rectification of uncalibrated pictures taken by the authors

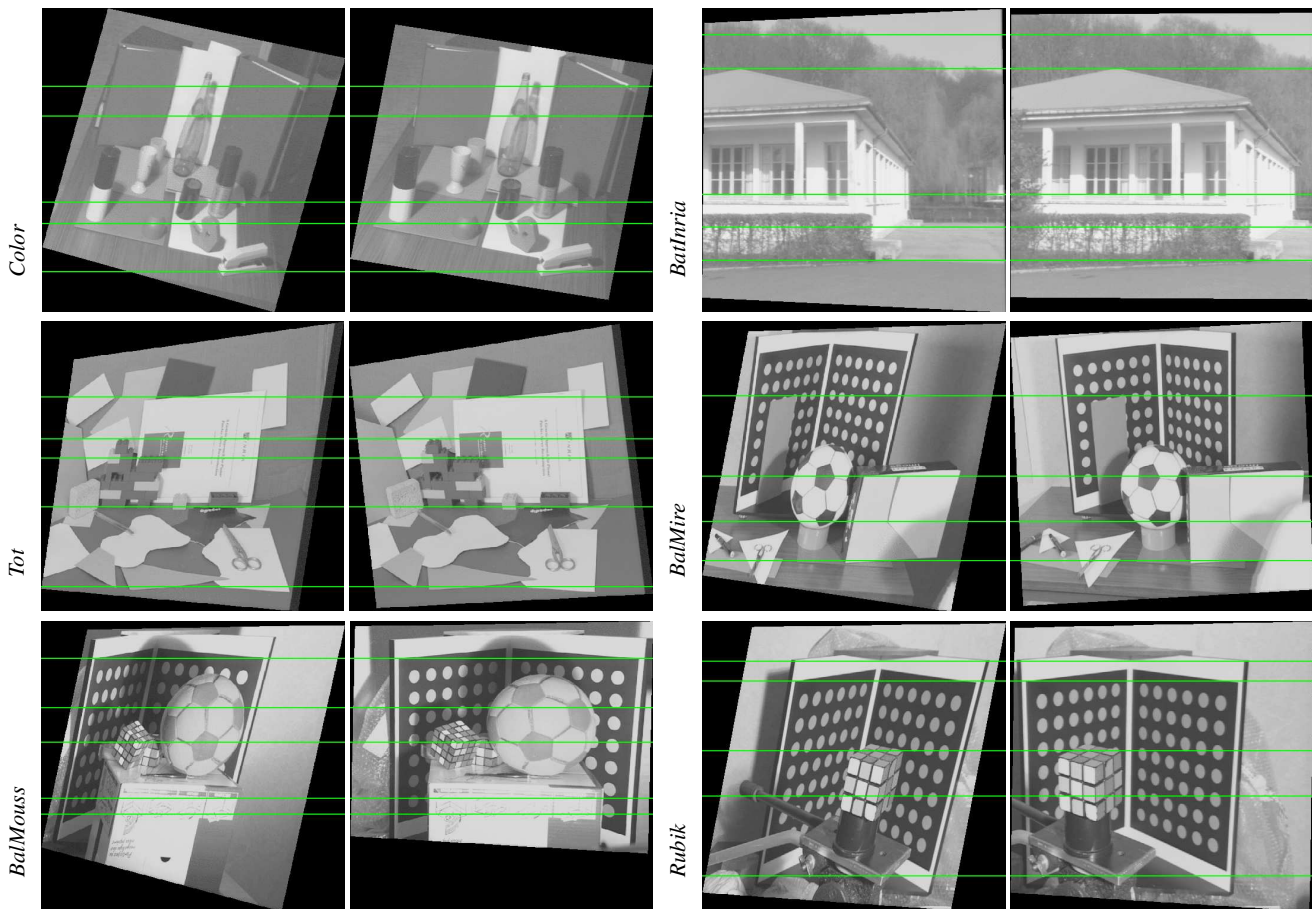


Fig. 3 SYNTIM rectified pairs from [11] (reproduced with permission from the authors).

Table 3 Rectication error and distortion metrics for the casual images.

Images	ϵ_{rec}	Ortho.	Asp. rt.
<i>Piazza Grande</i>	0.288	89.790	0.995
<i>Pozzoveggiani</i>	0.351	90.205	1.005
<i>Tribuna</i>	0.563	90.496	1.012

are reported in (Fig. 5). The good quality of these images allowed to completely automate the process: first SIFT are extracted in both images and descriptors are matched as reported in [14]; then outliers are discarded by fitting a fundamental matrix to the matching points with RANSAC. This matching, however, is not part of the rectification method, which indeed starts from a set of correct matches, however they have been obtained.

The reader might notice that these pairs present notable change of the point of view and, consequently, a significant verging angle. Table 3 reports rectication error and distortion metrics for these images.

The running time is mostly spent inside the Levenberg-Marquardt routine computing the cost function. It depends largely on the number of points being used: for example it

took 3s to rectify the “Sport” pair with 40 matches, and it took 15s to rectify “Piazza Grande” with 292 matches. The warping time (to be added), which instead depends on the *rectified* image size, is about 5s for “Sport” (768x576) and 3.5s for “Piazza Grande” (512x384). These figures refers to an unoptimized MATLAB implementation on a MacBook Pro, 2.4 GHz Intel Core 2 Duo.

5 Discussion

We presented a new method for the epipolar rectification of uncalibrated stereo pairs which approximates the Euclidean (calibrated) case by enforcing the rectifying transformation to be a collineation induced by the plane at infinity. The method is based on the minimization of a cost function that has only six degrees of freedom and does not need any specific initialization. The results are close to the target Euclidean rectification and compares favorably with state-of-the-art uncalibrated methods, in terms of distortion applied to the rectified images. Given the general utility of rectification, a Matlab toolkit is available on the web².

² <http://profs.sci.univr.it/~fusiello/demo/rect>



Fig. 4 VSG images [16]. Original pair (left) and quasi-Euclidean rectification (right).

Acknowledgments

The use of VLFeat by A. Vedaldi and B. Fulkerson, Matlab functions by P. Kovesi, SYNTIM (INRIA, Rocquencourt)

images³ and VSG (Vision Systems Group, Dublin City Uni-

³ <http://www-rocq.inria.fr/~tarell/syntim/paires.html>

versity) images⁴ is gratefully acknowledged. Thanks to F. Isgrò for granting permission to reproduce the images from his PhD thesis.

References

1. A. Al-Zahrani, S. S. Ipson, and J. G. B. Haigh. Applications of a direct algorithm for the rectification of uncalibrated images. *Information Sciences – Informatics and Computer Science*, 160(1-4):53–71, 2004.
2. A. Fusiello, E. Trucco, and A. Verri. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16–22, 2000.
3. Joshua Gluckman and Shree K. Nayar. Rectifying transformations that minimize resampling effects. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 01:111, 2001.
4. R. Hartley and R. Gupta. Computing matched-epipolar projections. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 549–555, New York, NY, June 15–17 1993.
5. R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd edition, 2003.
6. R. I. Hartley. Estimation of relative camera position for uncalibrated cameras. In *Proceedings of the European Conference on Computer Vision*, pages 579–587, Santa Margherita L., 1992.
7. R.I. Hartley. Theory and practice of projective rectification. *International Journal of Computer Vision*, 35(2):1–16, November 1999.
8. A. Heyden and M. Pollefeys. Multiple view geometry. In G. Medioni and S. B. Kang, editors, *Emerging Topics in Computer Vision*, pages 45 – 107. Prentice Hall, 2005.
9. Yu-Hua Yu Hsien-Huang P. Wu. Projective rectification with reduced geometric distortion for stereo vision and stereoscopic video. *Journal of Intelligent and Robotic Systems*, 42(1):Pages 71 – 94, Jan 2005.
10. M. Irani, P. Anandan, and S. Hsu. Mosaic based representations of video sequences and their applications. In *Proceedings of the International Conference on Computer Vision*, pages 605–611, 1995.
11. F. Isgrò and E. Trucco. Projective rectification without epipolar geometry. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages I:94–99, Fort Collins, CO, June 23–25 1999.
12. S. Laveau and O. Faugeras. 3-D scene representation as a collection of images. In *Proceedings of the International Conference on Pattern Recognition*, volume 1, pages 689–691, 1994.
13. C. Loop and Z. Zhang. Computing rectifying homographies for stereo vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages I:125–131, Fort Collins, CO, June 23–25 1999.
14. David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
15. Q.-T. Luong and O. D. Faugeras. The fundamental matrix: Theory, algorithms, and stability analysis. *International Journal of Computer Vision*, 17:43–75, 1996.
16. John Mallon and Paul F. Whelan. Projective rectification from the fundamental matrix. *Image and Vision Computing*, 23(7):643 – 650, 2005.
17. L. Robert, M. Buffa, and M. Hebert. Weakly-calibrated stereo perception for rover navigation. In *ICCV*, pages 46–51, 1995.
18. P. Sturm. On focal length calibration from two views. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume II, pages 145–150, Kauai, USA, 2001.

⁴ <http://elm.eng.dcu.ie/~vsl/vsgcode.html>



Fig. 5 Original uncalibrated pairs (left) and Quasi-Euclidean rectification (right) obtained without manual intervention.