ESTIMATION OF EPIPOLAR GEOMETRY VIA THE RADON TRANSFORM

*Stefan Lehmann, Andrew P. Bradley, and I. Vaughan L. Clarkson**

School of Information Technology and Electrical Engineering
The University of Queensland, 4072 AUSTRALIA
(lehmann, bradley, v.clarkson)@itee.uq.edu.au

ABSTRACT

One of the key problems in computer vision is the recovery of epipolar geometry constraints between different camera views. The majority of existing techniques rely on point correspondences, which are typically perturbed by mismatches and noise, hence limiting the accuracy of these techniques. To overcome these limitations, we propose a novel approach that estimates epipolar geometry constraints based on a statistical model in the RADON domain. The method requires no correspondences, explicit constraints on the data or assumptions regarding the scene structure. Results are presented on both synthetic and real data that show the method's robustness to noise and outliers.

1. INTRODUCTION

A basic step in scene analysis from multiple views is to recover the epipolar geometry of an imaging system. The fundamental matrix encapsulates the epipolar geometry and is used in many computer vision tasks, such as scene reconstruction [1], image rectification [1] and stereo matching [2]. Various methods have been proposed that estimate the fundamental matrix from point correspondences [1]. However, determining correspondences is still an open problem. In particular, matching algorithms often fail to generate correct correspondences due to large motions, occlusions or ambiguities.

Correspondence outliers are known to severely degrade the accuracy of the conventional methods used to estimate the fundamental matrix [3]. While robust methods such as *RANSAC* [1] are able to reduce these problems by rejecting outliers, they also eliminate valid features and therefore yield incomplete feature models for reconstructing 3D scenes [4].

Various efforts have been undertaken to avoid correspondence in Structure and Motion. The largest class among them, *direct methods* [5,6], are based on the *brightness change constraint equation* proposed in [7]. However, this constraint assumes small interframe motions, sufficient texturing of surfaces, and sufficiently slow variations in the lighting conditions [5]. Other techniques that avoid correspondences rely on scene decomposition, additional geometric constraints or assumptions about the image behavior [8–10], which typically restricts the generality of the approaches. In [4], the use of the Expectation-Maximization (EM) algorithm is proposed to iteratively estimate structure and motion without correspondences. It is acknowledged that the EM algorithm can converge to a local minimum.

In this paper, a novel approach is presented that estimates the epipolar geometry directly from two sets of features *without* any prior information on point correspondences and *without* enforcing explicit constraints on the data or making assumptions about the scene structure. The two sets of features are 2D orthographic projections of a set of 3D object features from different viewpoints. The estimation problem is statistically modelled in the RADON domain and leads to the maximization of a three-dimensional cost function. This cost function reflects the similarities between probability density functions in the RADON domains of the two feature spaces. A main advantage of this approach is shown to be its high robustness to noise and outliers in the feature data.

Section 2 outlines the principle of the proposed approach. An algorithmic overview will be proposed in Section 3. Section 4 presents an experimental evaluation on both synthetic and real data. Finally, conclusions are presented in Section 5.

2. STATISTICAL MODEL IN THE RADON DOMAIN

2.1. Epipolar Geometry in the Radon domain

Let us consider a $3D$ scene which is represented by a set S_3 of N $3D$ feature points. S_3 is then projected onto the viewing planes of two orthographic cameras in general positions, resulting in two sets of $2D$ feature points S_2 and S_2' respectively. We showed in [11] how the $3D/2D$ feature sets can be mathematically represented as superpositions of Dirac functions at the appropriate feature locations. The appropriate mathematical derivations were based on the theoretical assumption that all $3D$ feature points $P_3 \in S_3$ are visible in both camera views, *i.e.* all corresponding $2D$ feature points $P_2 \in S_2$ and $P_2' \in S_2'$. In our theoretical model, the locations of the $2D$ features in S_2 and S_2' are precise, i.e. unaffected by distortions such as noise and outliers. Under these ideal as-

*Vaughan Clarkson is on study leave throughout 2005 at the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC, V6T 1Z4, Canada

sumptions, each feature point $P_2 = (x, y)^T \in S_2$ must have a correspondence $P_2' = (x', y')^T \in S_2'$ that lies precisely on the corresponding ideal epipolar line in the second view. An analogous relationship holds in the reverse direction. Under orthographic projection, all epipolar lines are parallel [11] and can thus be denoted as:

$$\cos(\alpha')\, x' + \sin(\alpha')\, y' = \cos(\alpha)\, x + \sin(\alpha)\, y + \lambda, \quad (1)$$

where $(x, y)^T$ are the coordinates of the features in the first view, $(x', y')^T$ the corresponding feature coordinates in the second view, α and α' denote the angles perpendicular to the epipolar lines in the first and second view respectively and λ represents the displacement.

Thus, if we model S_2 by a suitable function $f_2(x, y)$ and S_2' by an analogous function $f_2'(x', y')$ respectively, the integrals of $f_2(x, y)$ along an epipolar line in the first view and $f_2'(x, y)$ along the corresponding epipolar line in the second view must be identical. This relationship follows from the energy conservation along corresponding epipolar lines. Following the notation in [12], it can be expressed by the following equation in the RADON domain:

$$\check{f}_2(p, \xi_\alpha) = \check{f}_2'(p + \lambda,\, \xi_{\alpha'}) \quad (2)$$

where $\check{f}_2(p, \xi_\alpha)$ and $\check{f}_2'(p, \xi_{\alpha'})$ denote the RADON transforms of $f_2(x, y)$ and $f_2'(x, y)$ and ξ_α and $\xi_{\alpha'}$ represent unit vectors at angles α and α' respectively.

We showed in [11] how $f_2(x, y)$ and $f_2'(x', y')$ can be expressed as superpositions of Dirac functions at the appropriate feature locations. In this case, the constraint in (2) would be met under the ideal assumptions stated above. In practice however, feature location data is inevitably subject to distortions such as noise and outliers which led us to the introduction of a statistical model.

2.2. Statistically modelling feature location uncertainty

Let us assume two $2D$ sets of N (non-ideal) feature points S_2 and S_2' corresponding to the first and second camera view respectively. We model S_2 and S_2' with the functions $f_2(x, y)$ and $f_2'(x, y)$ respectively, where $f_2(x, y)$ and $f_2'(x, y)$ are defined as the superpositions of $2D$ Gaussian functions scaled by the factor $1/N$. The $2D$ Gaussians reflect statistical independence in x- and y- coordinates, share a common standard deviation σ and are centered at the appropriate feature locations:

$$f_2(x, y) = \frac{1}{2\pi\sigma^2 N} \sum_{k=1}^{N} e^{-\frac{1}{2\sigma^2}\left[(x-x_k)^2+(y-y_k)^2\right]} \quad (3)$$

$$f_2'(x, y) = \frac{1}{2\pi\sigma^2 N} \sum_{k=1}^{N} e^{-\frac{1}{2\sigma^2}\left[(x-x_k')^2+(y-y_k')^2\right]}, \quad (4)$$

where (x_k, y_k) and (x_k', y_k') are the feature coordinates in the first and second camera view respectively. It follows from (3) and (4) that

$$\iint_{-\infty}^{\infty} f_2(x, y)\, \mathrm{d}x\, \mathrm{d}y = \iint_{-\infty}^{\infty} f_2'(x, y)\, \mathrm{d}x\, \mathrm{d}y = 1. \quad (5)$$

The functions $f_2(x, y)$ and $f_2'(x, y)$ can be interpreted as probability density functions (*PDFs*), representing the likelihood that any individual location (x, y) is statistically "covered" by the entity of all feature points of the respective image under a Gaussian noise model.

Now, evaluating the RADON transforms $\check{f}_2(p, \xi_{\hat{\alpha}})$ and $\check{f}_2'(p, \xi_{\hat{\alpha}'})$ of $f_2(x, y)$ and $f_2'(x, y)$ at arbitrary angles $\hat{\alpha}$ and $\hat{\alpha}'$ eliminates one of the two dimensions of the underlying Gaussian kernel and therefore results in a superposition of $1D$ Gaussian functions. This effect is due to the circularly symmetric nature of the $2D$-Gaussian function and is depicted in Figure 1 for the first camera view (S_2). In this Figure, the circles represent the Gaussian kernels centered at the feature locations ($S_2 = \{P_1, P_2, P_3\}$ in this example). According to (3), $f_2(x, y)$ is the superposition of these Gaussians. Thus, $\check{f}_2(p, \xi_{\hat{\alpha}})$ and $\check{f}_2'(p, \xi_{\hat{\alpha}'})$ can be directly evaluated by first projecting the (distorted) locations of the features in S_2 and S_2' onto the lines intersecting the origins at angles $\hat{\alpha}$ and $\hat{\alpha}'$ respectively and subsequently calculating the superpositions of $1D$-Gaussians centered at the projected feature locations (Figure 1).

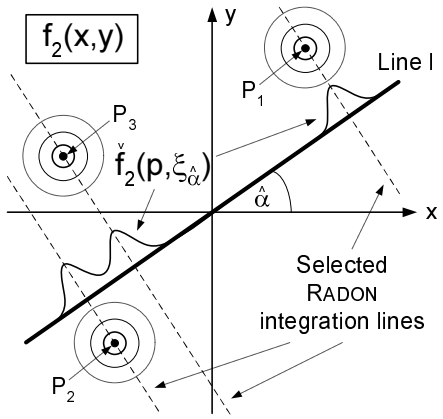

Fig. 1. Probability density function $f_2(x, y)$ and RADON transform $\check{f}_2(p, \xi_{\hat{\alpha}})$ for a set of three features. The RADON integration lines run perpendicular to the line l that intersects the origin at angle $\hat{\alpha}$.

The resulting RADON transforms implicitly bear the following statistical interpretation: Let us consider the two RADON transforms $\check{f}_2(p, \xi_{\hat{\alpha}})$ and $\check{f}_2'(p, \xi_{\hat{\alpha}'})$ at arbitrary but fixed angles $\hat{\alpha}$ and $\hat{\alpha}'$, resulting in the two $1D$ functions $\check{f}_2(p)$ and $\check{f}_2'(p)$. Let us further define two lines l and l' that run through the origin of the coordinate system at angles $\hat{\alpha}$ and $\hat{\alpha}'$ respectively. The functions $\check{f}_2(p)$ and $\check{f}_2'(p)$ express the statistical probabilities $q(p)$ and $q'(p)$ that any point P that lies on l or l' respectively at distance p from the origin results from the projection of the respective $2D$ feature entity (S_2 or

S_2') onto this line. This relationship can be derived based on conditional probabilities.

2.3. Cost function

Despite the practical presence of distortions in the feature data, the similarity between the two RADON transforms $\check{f}_2(p, \xi_{\hat{\alpha}})$ and $\check{f}_2'(p + \lambda, \xi_{\hat{\alpha}'})$ is expected to reach its maximum when the two functions are evaluated in the proximity of the parameter triplet $(\alpha, \alpha', \lambda)$ that satisfies the epipolar constraint in (2). Thus, the extraction of the epipolar geometry comes down to the maximization of a suitable similarity measure in a three-dimensional parameter space. In the current implementation, the cost function is the maximum of the normalized cross-correlation between $\check{f}_2(p, \xi_{\hat{\alpha}})$ and $\check{f}_2'(p, \xi_{\hat{\alpha}'})$ in the p-dimension. This cost function is expected to reach its maximum value at $\hat{\alpha} = \alpha$ and $\hat{\alpha}' = \alpha'$, *i.e.* when the maximum similarity is expected between the two corresponding $1D$ *PDFs* $q(p)$ and $q'(p)$.

3. ALGORITHMIC OVERVIEW

The algorithm proceeds as follows:

1. A coarse search is performed over the cost function described in Section 2.3 for a set of discrete angles α_k, α_k'. This search yields the angle pair $(\hat{\alpha}_s, \hat{\alpha}_s')$.

2. Using $(\hat{\alpha}_s, \hat{\alpha}_s')$ as the initial values, a Levenberg-Marquardt search over the cost function yields the final estimates $(\hat{\alpha}, \hat{\alpha}')$.

3. The index at which the normalized cross-correlation in the cost function at angles $(\hat{\alpha}, \hat{\alpha}')$ reaches its maximum determines the final estimate of the displacement $\hat{\lambda}$.

4. Based on the estimates $(\hat{\alpha}, \hat{\alpha}', \hat{\lambda})$, the epipolar geometry can be recovered using (1).

4. EXPERIMENTAL RESULTS

4.1. Experiments on synthetic data

Multiple sets S_3 of $3D$-features were synthetically generated, where the feature number $N = 100$. The feature points of S_3 were subject to three different types of $3D$ scene transformations to generate the appropriate feature sets S_2 and S_2'. The transformations are characterized by the azimuth ϕ (rotation about y-axis), the elevation θ (rotation about x-axis) and the translations x_0 and y_0 in x- and y-directions respectively:

- Type 1: $\phi = \ 5°$, $\theta = \ 2°$, $x_0 = \ 7$, $y_0 = \ 8$;
- Type 2: $\phi = 20°$, $\theta = 10°$, $x_0 = 10$, $y_0 = 10$;
- Type 3: $\phi = 45°$, $\theta = 20°$, $x_0 = 15$, $y_0 = 10$.

For each transformation type, various percentages of feature outliers were generated ($[0, 1, 2, 4, 8, 16, 32, 48]\%$), *i.e.* an appropriate number of features of S_2' were randomly displaced within a maximum disparity range according to a uniform distribution. To additionally simulate location noise, the coordinates of all features of S_2 and S_2' were rounded to integer values. For each transformation type and outlier percentage, 100 independent tests were performed to ensure the statistical reliability of the results. In each test, new feature sets S_2 and S_2' were synthetically generated and passed to both *RANSAC* [1] and the proposed Statistical RADON Approach (SRA). Performance measures were obtained for both methods by comparing the estimated and ideal fundamental matrices based on a scheme proposed in [3]. These measures represent the Mean Distances (MDS) in pixels between points on the ideal and estimated epipolar lines. Finally, the MDS were averaged over the 100 tests. Figure 2 depicts the averaged MDS of both estimation methods for all three transformation types against the outlier percentage. Due to the

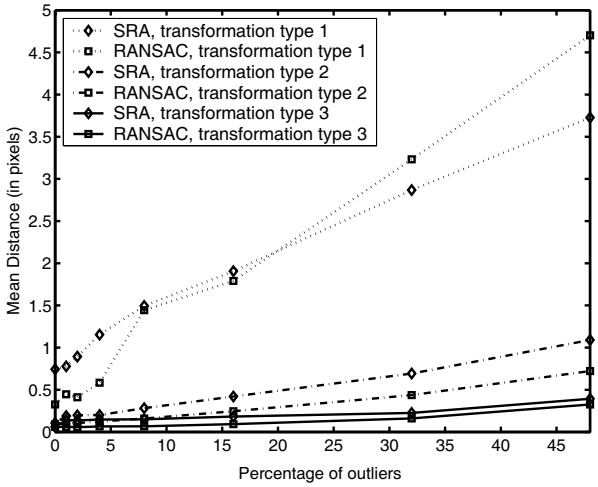

Fig. 2. Trends of the estimation errors of RANSAC and the proposed SRA based on synthetically generated features.

lack of a feature matching step during the generation of the data, the evaluation is biased in favour of RANSAC. The aim of the synthetic evaluation is to investigate the trends of the methods with respect to the outlier percentage. It can be seen from Figure 2 that both estimation methods show very similar trends and the discrepancy between the MDS is small (< 1 pixel). For small transformations and high outlier percentages, the proposed approach showed even smaller MDS than RANSAC. This confirms the high robustness of the proposed approach with respect to noise and outliers.

4.2. Experiments on real images

Features were automatically extracted with a HARRIS corner detector from neighbouring camera images of an image sequence used previously [4]. The eleven images of the se-

quence were taken of a rotating cube with a telephoto lens and hence can be considered as orthographic projections. For each of the ten neighbouring image pairs, MDs were computed for the Linear Raw Method (LRM), *RANSAC* [1] and the proposed Statistical RADON Approach (SRA) as described in Section 4.1. LRM is an adjustment of the 8-point algorithm [1] to the orthographic projection case. The MDs were averaged over all image pairs. Figure 3 shows both the means and the standard deviations of the MDs. It can be seen from Figure 3 that the proposed approach yields the smallest estimation error, despite its advantage of not requiring correspondence data. Figure 4 depicts neighbouring cube images from the sequence. Since the motion within the sequence is a rotation about the vertical axis, all epipolar lines should be horizontal. It can be seen from Figure 4 that the epipolar lines from SRA are more accurate than the lines from RANSAC.

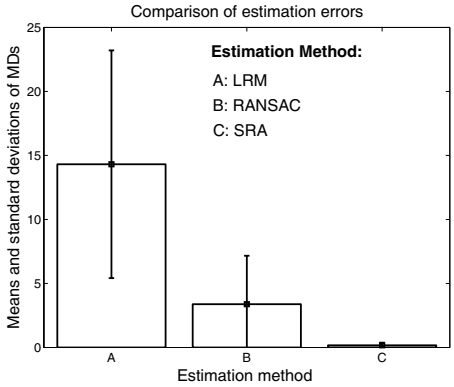

Fig. 3. Means and standard deviations of the MDs for three estimation methods based on ten cube image pairs.

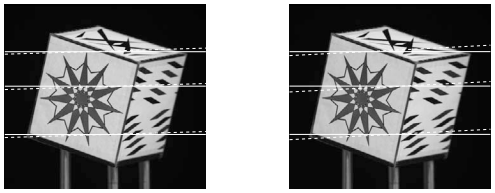

Fig. 4. Cube image pair with epipolar lines from RANSAC (dashed lines) and the proposed SRA (solid lines).

5. CONCLUSIONS AND FUTURE WORK

In this paper, a statistical approach was presented to recover robustly the epipolar geometry from two sets of sparse features under an orthographic camera model. The method operates in the RADON domain and does not require any correspondence information. This constitutes a major advantage over conventional methods, since establishing correspondences is difficult in practice. The method showed to be extremely robust towards noise and outliers in the feature data.

One of the future aims is to further increase the overall robustness by automatically detecting and discarding outliers. Additional goals include the use of this method for feature matching and its extension to perspective projections.

6. REFERENCES

[1] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.

[2] Z. Zhang, R. Deriche, O. Faugeras, and Q. T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," *Artificial Intelligence Journal*, vol. 78, no. 1-2, pp. 87–119, 1994.

[3] Z. Zhang, "Determining the epipolar geometry and its uncertainty: A review," *Int. J. Computer Vision*, vol. 27, no. 2, pp. 161–198, 1998.

[4] F. Dellaert, S. M. Seitz, C. E. Thorpe, and S. Thrun, "Structure from motion without correspondences," in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR'00)*, June 2000, vol. 2, pp. 557–564.

[5] S. Negahdaripour and B. K. P. Horn, "A direct method for locating the focus of expansion," *Comput. Vision Graph. Image Process.*, vol. 46, no. 3, pp. 303–326, 1989.

[6] G. P. Stein and A. Shashua, "Model-based brightness constraints: On direct estimation of structure and motion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 9, pp. 992–1015, 2000.

[7] S. Negahdaripour and B. K. P. Horn, "Direct passive navigation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 9, no. 1, pp. 168–176, 1987.

[8] M. Irani, "Multi-frame optical flow estimation using subspace constraints," in *Proc. IEEE Int. Conf. on Computer Vision (ICCV'99)*, Sept. 1999, pp. 626–633.

[9] S. Roy and I. J. Cox, "Motion without structure," in *Proc. IEEE Int. Conf. on Pattern Recognition (ICPR'96)*, Aug. 1996, vol. 1, pp. 728–734.

[10] A. J. Yezzi and S. Soatto, "Structure from motion for scenes without features," in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR'03)*, 2003, vol. 1, pp. 525–532.

[11] S. Lehmann, I. V. L. Clarkson, A. P. Bradley, J. Williams, and P. J. Kootsookos, "Robust fundamental matrix estimation without correspondences," in *APRS Workshop on Digital Image Computing (WDIC 2005)*, Feb. 2005, vol. 1, pp. 97–102.

[12] S. R. Deans, *The Radon transform and some of its applications*, Wiley, 1983.