Fundamental Matrix Estimation via TIP - Transfer of Invariant Parameters

Frank Riggi, Matthew Toews and Tal Arbel
Centre for Intelligent Machines, McGill University
Montreal, Quebec, Canada
{friggi, mtoews, arbel}@cim.mcgill.ca

Abstract

The fundamental matrix (FM) represents the perspective transform between two or more uncalibrated images of a stationary scene, and is traditionally estimated based on 2-parameter point-to-point correspondences between image pairs. Recent invariant correspondence techniques however, provide robust correspondences in terms of 4 to 6-parameter invariant regions. Such correspondences contain important information regarding scene geometry, information which is lost in FM estimation techniques based solely on 2-parameter point translation. In this article, we present a method of incorporating this additional information into point-based FM estimation routines, entitled TIP (transfer of invariant parameters). The TIP method transforms invariant correspondence parameters into additional point correspondences, which can be used with FM estimation routines. Experimentation shows that the TIP methods result in more robust FM estimates in the case of sparse correspondence, and allows estimation based on as few as 3 correspondences in the case of affine-invariant features.

1. Introduction

Consider a set of 2D images of a scene acquired from a moving camera. Should the epipolar geometry be known, any 3D real world point that is captured in one image can be mapped to a line in each subsequent image. With constraints provided by point-to-point correspondences, the three dimensional world can be recovered from a set of 2D images. In an uncalibrated system, the epipolar geometry is estimated by the fundamental matrix (FM), a 3×3 matrix of rank two, estimated by using the corresponding matches from one image to the next. The accuracy of the FM estimate (and subsequently the scene geometry) is therefore crucial for applications such as 3D reconstruction and pose estimation.

The accuracy of the FM estimate is a function of the number of correct correspondences attained between images. Most algorithms rely on a minimum of eight such correspondences, i.e. the 8-point algorithm [8, 4], although other techniques can estimate the epipolar geometry with fewer correspondences such as the seven, six and five point algorithms [3, 13, 14]. FM estimation can be based on other types of correspondences such as conics and curves [7, 1], however. Recently, *invariant features* have been shown to provide fast and robust correspondence over geometrical deformations such scale [9, 10] and affine [17, 11] transformations, in addition to illumination changes. As a result, such features have proven useful in estimating the FM over large changes in viewpoint [2]. When combined with robust sampling techniques, accurate FM estimates can be obtained, provided a sufficient number of correct correspondences exist. However there are many circumstances in which a sufficient number of correct correspondences cannot be determined, adversely effecting the accuracy of fundamental matrix estimation. This begs the question: Can the additional parameters of invariant features indeed be incorporated in FM estimation?

Our contribution in this paper is a method of incorporating additional information from invariant feature correspondences into existing point-based FM estimation routines, which we refer to as TIP (Transfer of Invariant Parameters). In the TIP method, geometrical parameters inherent to invariant features are converted into additional point-correspondences in order to increase the robustness of FM estimation, particularly in cases where tradition estimation fails. This method (1) leads to an increased number of robust feature points for correspondence, (2) reduces the minimum number of correspondences required for FM estimation and (3) is contingent on the number of parameters associated with a particular invariant feature. In Section 2, we discuss the uses of invariant features in FM estimation, in Section 3 we then introduce the TIP method and in Section 4 we outline our experiments and present our results, demonstrating the increased robustness presented by our technique over traditional FM computations using the Harris-Affine detector of Mikolajczyk and Schmid [11].

2. Invariant features and FM estimation

The majority of techniques aimed at estimating the fundamental matrix are based on point-to-point correspondences between images. Consequently, a large body of literature is devoted to techniques that automatically determine such correspondences between images [9, 10, 11, 15]. In

0-7695-2521-0/06/$20.00 (c) 2006 IEEE

general terms, the problem of finding correspondences can be thought of as determining relations between geometrical structures in different images.

Early approaches to feature-based correspondence attempted to determine a 2-parameter $\{x, y\}$ feature displacement from one image to the next. Such approaches are said to be translation invariant, as such features are matchable in the presence of $\{x, y\}$ image translation. It was subsequently observed that the size of the pixel window was intimately related to the pattern being matched, resulting in scale-invariant feature detectors that focused on extracting and matching 3-parameter regions $\{x, y, \sigma\}$ [15] in the presence of image scale change σ in addition to translation. Further work has extended invariance to full 4-parameter similarity [9, 10] and 5-parameter [1] affine [11] transformation.

Traditionally, the fundamental matrix is estimated with a minimum of 8 point-to-point (2-parameter $\{x, y\}$) correspondences. Since the FM (i.e. F) has seven degrees of freedom and is constrained such that $det(F) = 0$, it can even be estimated with as few as 7 point-to-point correspondences [3]. While these methods are popular and efficient, correspondences are limited to simple 2-parameter point translations. When using invariant feature correspondences and subsequently basing FM estimation on only 2-parameter translations, important high order information about scene geometry is lost. Special-purpose estimation techniques have been developed for specific higher-order correspondence parameterizations (i.e. conics [7]), but have not become widely used since these methods typically require the solution of a high order polynomial. In this paper, we wish to exploit additional information provided by invariant feature correspondences using widely-used point-based FM estimation techniques. By the transfer of these invariant parameters discussed in the following section, the total number of matches required for a stable FM estimate can be reduced over 2-parameter $\{x, y\}$ point-to-point correspondences.

3. TIP - Transfer of Invariant Parameters

We propose TIP as a method to incorporate local geometry of each invariant feature correspondence in a simple yet effective manner while still being able to use standard point-based methods of FM estimation. Although generally applicable to a variety of invariant correspondence techniques, we describe TIP in the context of affine-invariant correspondence, where each correspondence represents a 6-parameter affine transform of a region from one image to the next.

For each correct correspondence, the corresponding coordinate pair is known (i.e. $(x, y) \longleftrightarrow (x', y')$). By extension, these two matched points share the same affine region. In the case of [11], this is determined in the second moment matrix and is represented by a fitted ellipse. Henceforth, we have knowledge of the elliptical geometric region surrounding the interest point in each image. We can therefore state that not only do we know the translation from one image to the next, but the transformation of one ellipse to the other $(\xi) \longleftrightarrow (\xi')$. By extension, we can surmise that any singular point or collection of points in or on the border of the ellipse are also transformed from one image to another given the known transform.

Once we are given the affine parameters of the major axis, minor axis and elliptical rotation $(\ell_1, \ell_2, \alpha) \longleftrightarrow (\ell_1', \ell_2', \alpha')$ as well as a dominant gradient orientation, how can we embed this additional information into an FM estimation algorithm in order to increase its accuracy? The premise of TIP is that the information contained in the additional invariant parameters provided can be transferred between both images through additional *child* point correspondences. As a result, TIP (1) reduces the original number of correspondences required (the *parent* points) and (2) increases the robustness of the FM estimate. Although we consider the affine case, generally the potential information contained in one invariant feature correspondence is a function of the number of parameters that the invariant feature provides, which in turn reduces the minimum number of correspondences required (refer to Table 1).

Table 1. Invariant feature information

Transform	No. Param	Parameters	Points Req'd
Translation	2	(x, y)	8
Similarity	4	(x, y, σ, θ)	4
Affine	6	$(x, y, \theta, \ell_1, \ell_2, \alpha)$	3

3.1. The affine case

For each correspondence in the affine case, we place *child* points along the border of the ellipse bounding the affine region. Since angles are not preserved in affine transforms, each elliptical affine region must be normalized to a unit circle. To accomplish this, an ellipse can be decomposed into the following components of $\xi = \Phi\ S\ \Phi^\top$ as indicated in Figure 1. Hence ξ must undergo a rotation of Φ^{-1} and stretch/squeeze of S^{-1}. Points are then placed in relation to the dominant orientation of the corresponding ellipse. If the dominant orientation θ of the feature point is unknown, it can be determined here as a sub-step by assigning an orientation through local gradients as demonstrated by Lowe [9].

Following the transform, we subsequently add a child point along the contour of the unit circle at the dominant orientation. To best represent an affine transform to the FM estimation algorithm, it would be beneficial to actually include two child points spaced $\pi/2$ radians apart. Ideally this would fit our requirement and each affine ellipse would be

[1] The affine transform of [11] is natively a 5 rather than 6-parameter transform since the affine matrix is symmetric. By examining local image gradients a 6th parameter can be found in the dominant orientation.

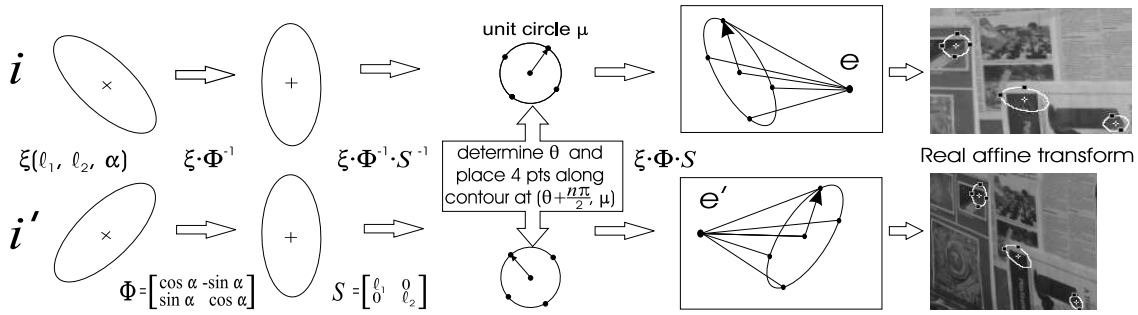

Figure 1. Determining child points between corresponding affine regions: (1) elliptical regions are transformed to unit circle μ (2) child points are selected along μ wrt θ (3) child points are re-projected to elliptical regions.

characterized by 3 points. However, in the event that an epipolar line crosses both the parent and child point, the information potential of that child point is diminished. Therefore, we suggest adding a sum of 4 child points (canonically placed $\pi/2r$ apart from θ) so that each affine correspondence is characterized by 5 point-to-point matches even though two are likely be redundant in most circumstances. The unit circle is then transformed back to an ellipse and the two matched feature points can now be characterized with five matched points. The collection of points can then be added to the *normalized 8-point algorithm* or some robust estimator in order to estimate the FM.

4. Experimentation

The goal of experimentation is to demonstrate that TIP can both (1) add robustness to point-based FM estimates in the case of sparse correspondence and (2) allow FM estimation for as low as 3 correspondences. Our methodology is to perform FM estimation between pairs of images of the same scene taken from different viewpoints. Since the affine invariant detector of Mikolajczyk and Schmid [11] performs best on patterns arising from planar surfaces, we test our method on a scene of planar surfaces containing a significant amount of image texture (see Figure 2 for examples).

Testing proceeds as follows: we sample a set of n affine correspondences (inliers) between an image pair. For each correspondence set, two FMs are estimated: F and F^{TIP}. F is estimated solely based on the sampled (*parent*) points for $n = \{20, \ldots, 7\}$ using the 8-point algorithm, except in the case of $n = 7$ where the 7-point algorithm must be used. F^{TIP} is estimated based on TIP *parent-child* points for $n = \{20, \ldots, 3\}$ using the 8-point algorithm. Note that F cannot be estimated for $n = \{6, \ldots, 3\}$, as a minimum of 7 point correspondences are required. For each n, 100 randomly generated sample sets are used. The implementations for the 7 and 8-point algorithms are those of OpenCV [6].

To compare our FM estimates (F and F^{TIP}), we evaluate each by projecting a set of pre-determined hand-labeled ground truth point correspondences which are dispersed throughout the image pair. Our error measure ε is the residual error of [5] and is defined as:

$$\varepsilon^2 = \frac{1}{N}\sum_i d(\mathbf{x}'_i, \mathrm{F}\mathbf{x}_i)^2 + d(\mathbf{x}_i, \mathrm{F}^\top \mathbf{x}_i)^2 \qquad (1)$$

where $[\mathbf{x}_i, \mathbf{x}'_i]$ are the known ground truth point coordinates and $l_i = \mathrm{F}^\top \mathbf{x}_i$ is the corresponding epipolar line in Image 1, that is projected from point $\mathbf{x}'_i$ in image 2. The distance from each point to its projected line is hence $d(x_i, l_i)$. It is squared and added to the error in the second image. So not to skew our analysis, in the event that ε is far beyond an expected normal error tolerance (due mostly to sampled degenerate point configurations), we shall count it as a *catastrophic* instance.

Figure 2 displays our results based on 3 image pairs where a single camera moves throughout a rigid scene. For each pair, the mean error was determined for all correspondences (top plot) and those considered non-*catastrophic* (middle plot). Included also is a histogram of catastrophic instances for each pair (bottom plot). Matches were determined using the SIFT descriptor and a thresholding technique similar to [9] and a backwards-forwards constraint. In order to properly compare the estimates, all remaining outlier correspondences were removed by hand. Further, we chose our catastrophic threshold to apply to those instances when $\varepsilon > 25$, whereupon by visual inspection ground truth points and their projected epipolar lines transferred by each FM estimate were considered extremely inaccurate.

When abundant correspondences exist (i.e. $n > 12$), the mean residual error of both methods are similar and catastrophic instances are virtually non-existent. However, as the number of correspondences decrease, the effectiveness of TIP becomes apparent as both the number of *catastrophic* instances and residual error are lower.

As is further observed, it is in fact possible to achieve FM estimates with fewer than 7 point correspondences by using TIP's *parent-child* point correspondences in conjunction with the 8-point algorithm. Attempting to determine an accurate FM with fewer correspondences is likely to be more difficult due to degenerate point configurations and the nature of the fitted ellipse. Nonetheless, determining an FM based on fewer correspondences will have important implications such as reducing the number of correspondences required in robust sampling and estimation tech-

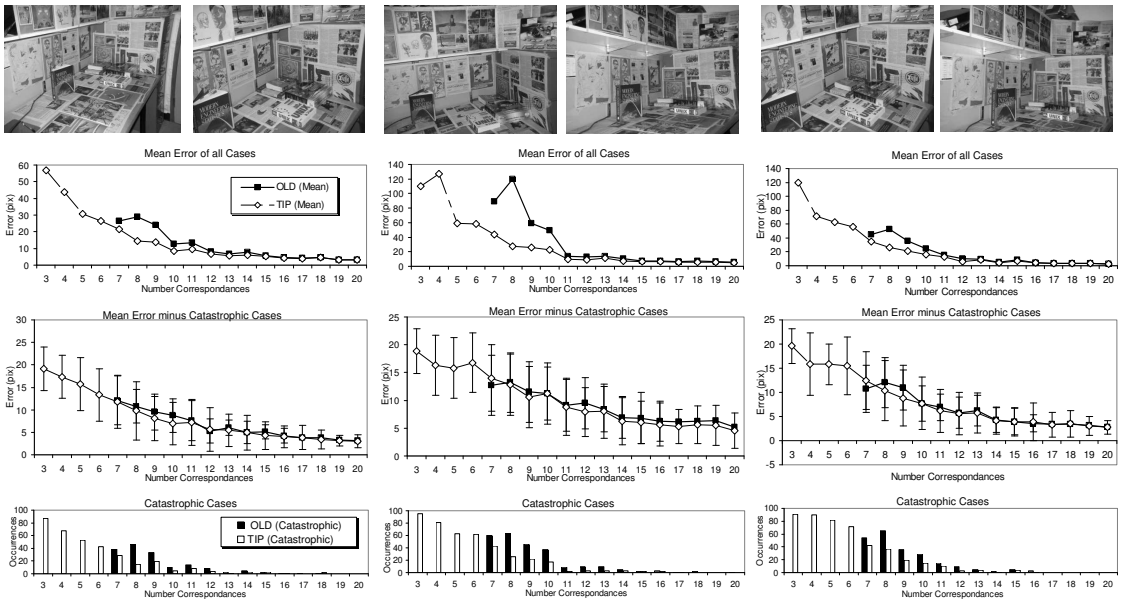

Figure 2. The three image pairs tested (top) and results: the mean error of all cases (plot 1); the mean error of non-catastrophic cases (plot 2) with stdev; and occurrences of catastrophic cases (plot 3).

niques (i.e. RANSAC) which in turn can reduce the number of trials [16] required.

While an estimate can be improved through TIP, we have observed circumstances in which estimate accuracy decreases slightly, mostly when sufficient correspondences are present. This can be attributed to the uncertainty associated with the fitted ellipses. While the affine invariant detector's ellipses are well suited for matching over affine transforms, they don't always fully model the geometric structure accurately and hence some are slightly off in relation to one another. We suspect that a pre-processing step of aligning the already matched ellipses would increase accuracy in these circumstances. Although testing involved planar scenes, application to non-planar scenes would be limited only by the accuracy of the detector [12].

5. Conclusions

In this paper, we discussed how TIP incorporates the information encoded within each affine invariant feature correspondence and include that information in fundamental matrix estimation. We introduced a method that increases robustness of FM estimates and demonstrated the potential to estimate the FM with as few as three affine invariant correspondences. In this paper we limited most of our discussion to affine invariant features but the TIP method introduced can be extended to other invariant features that further model geometric shape rather than simple $\{x, y\}$ translations.

References

[1] R. Berthilsson and K. Åström. Reconstruction of 3D-curves from 2D-images using affine shape methods for curves. In *CVPR*, 1997.

[2] M. Brown and D. G. Lowe. Invariant features from interest point groups. In *BMVC*, 2002.

[3] R. I. Hartley. Projective reconstruction and invariants from multiple images. *IEEE Trans. PAMI*, 16(10):1036–1041, 1994.

[4] R. I. Hartley. In defense of the 8-point algorithm. *IEEE Trans. PAMI*, 19(6):1064–1070, June 1997.

[5] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN:0521540518, second edition, 2004.

[6] Intel. Open source computer vision library. http://www.intel.com/technology/computing/opencv/.

[7] F. Kahl and A. Heyden. Using conic correspondence in two images to estimate the epipolar geometry. In *ICCV*, 1998.

[8] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, September 1981.

[9] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[10] K. Mikolajczyk and C. Schmid. Indexing based on scale invariant interest points. In *ICCV*, 2001.

[11] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *IJCV*, 60(1):63, 2004.

[12] P. Moreels and P. Perona. Evaluation of features detectors and descriptors based on 3d objects. In *Proc of ICCV*, 2005.

[13] D. Nistér. An efficient solution to the five-point relative pose problem. *IEEE Trans. PAMI*, 26(6):756–770, 2004.

[14] L. Quan. Invariants of six points and projective reconstruction from three uncalibrated images. *IEEE Trans. PAMI*, 17(1):34 – 46, Jan 1995.

[15] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Trans. PAMI*, 19(5):530534, 1997.

[16] R. Szeliski. Image alignment and stitching: A tutorial. draft, Microsoft Corp., 2005.

[17] T. Tuytelaars and L. V. Gool. Wide baseline stereo based on local, affinely invariant regions. In *BMVC*, 2000.