On the Direct Estimation of the Fundamental Matrix

Yaser Sheikh
Robotics Institute
Carnegie Mellon University
yaser@cs.cmu.edu

Asaad Hakeem
Computer Vision Laboratory
University of Central Florida
ahakeem@cs.ucf.edu

Mubarak Shah
Computer Vision Laboratory
University of Central Florida
shah@cs.ucf.edu

Abstract

The fundamental matrix is a central construct in the analysis of images captured from a pair of cameras and many feature-based methods have been proposed for its computation. In this paper, we propose a direct *method for estimating the fundamental matrix where the motion between the frames is small (e.g. between successive frames of a video). To achieve this, a warping function is presented for the fundamental matrix by using the brightness constancy constraint in conjunction with geometric constraints. Using this warping function, an iterative hierarchical algorithm is described to recover accurate estimates of the fundamental matrix. We present results of experimentation to evaluate the performance of the proposed approach and demonstrate improved accuracy in the computation of the fundamental matrix.*

1. Introduction

Traditionally, there have been two paradigms for estimating motion parameters between images. The first paradigm, called the *direct approach*, recovers motion parameters directly from image measurables (such as gradients or intensities) at each pixel in the image. Direct algorithms such as [8, 2, 15, 10, 13, 18, 17] have been proposed that provide highly accurate and stable results for many motion models. The second paradigm, called the *feature-based approach*, advocates the use of salient features or interest points, concentrating on the areas of the image where successful correspondence can be expected. A wide variety of algorithms have been proposed under the umbrella of the feature-based paradigm, for the estimation of virtually all known models of motion. A comprehensive survey can be found in [7]. The cases for these paradigms have been espoused by Irani and Anandan in [10] and Torr and Zisserman in [19], for direct approaches and feature-based approaches respectively. While the debate has brewed over the merits and demerits of feature-based methods and direct approaches for the estimation of various motion models (such as the affine transformation, the planar projective transformation and the trifocal tensor), this debate has been moot for the estimation of the fundamental matrix since only feature-based estimation approaches exist for its computation.

The fundamental matrix, itself, is a central construct in the analysis of images captured from a pair of uncalibrated cameras. It has been used widely in many different applications including object and action recognition, multi-camera tracking, view-interpolation and structure from motion. Direct methods typically employ the brightness constancy constraint [8], which is often approximated by assuming small motion and linearizing around the point of estimation, allowing computation of the normal component of the flow at each pixel. However, it is known that the fundamental matrix cannot be computed from normal flow alone. It is known that there are, in general, no geometric constraints on point to line correspondences between a pair of views, since a point in an image charts out a line in 3D, and a line in an image (from the normal flow) defines a plane in 3D and in general every line and plane intersect. This has been previously described by Stein and Shashua in [17]. However, by defining a new transformation of images, referred to as *pseudo-warping*, we show that given an initial estimate, highly accurate computation of the fundamental matrix is, in fact, possible within the paradigm of direct estimation. In the next section we place this work in context of earlier research both in terms of fundamental matrix computation, and within the collection of direct methods. In Section 2 we define pseudo-warping and derive the transformation functions for the fundamental matrix, followed by a description of the minimization scheme for estimating the fundamental matrix in Section 3. In Section 4 we compare the proposed approach with standard estimation algorithms and demonstrate cases where it out-performs these existing methods. We conclude in Section 5 with final thoughts and a summary of the proposed approach.

1.1. Preceding Work

Longuet-Higgins introduced the essential matrix in his seminal article [14] and proposed the eight-point algorithm

1-4244-1180-7/07/$25.00 ©2007 IEEE

for its computation. The fundamental matrix, the analogue of the essential matrix for uncalibrated cameras, was later simultaneously proposed by Faugeras in [4] and Hartley in [5]. Since then a multitude of approaches have emerged for its estimation, that can be categorized as linear methods and iterative methods. The classic eight-point algorithm is a widely used linear approach to estimating the fundamental matrix, usually in the 'normalized' form recommended by Hartley in [6]. Linear methods are quick, easy to code and usually provide good estimates in the absence of outliers and they are often used to initialize iterative methods that are more accurate. Iterative methods minimize some symmetric distance, such as the distance between points and epipolar lines (as defined by a candidate fundamental matrix). If an appropriate noise model is selected, an algorithm to find the maximum likelihood estimate of $\mathbf{F}$ can be readily devised by minimizing the reprojection error,

$$\sum_i d(\mathbf{x}_i, \bar{\mathbf{x}}_i)^2 + d(\mathbf{x}'_i, \bar{\mathbf{x}}'_i)^2 \tag{1}$$

where $(\mathbf{x}_i, \mathbf{x}'_i)$ are the measured correspondences, and $(\bar{\mathbf{x}}_i, \bar{\mathbf{x}}'_i)$ are the estimated correspondences that exactly satisfy $\bar{\mathbf{x}}'^T \mathbf{F} \bar{\mathbf{x}} = 0$ for the estimated (singular) matrix $\mathbf{F}$. Non-linear techniques such as Levenberg-Marquardt or Newton-Raphson are used to minimize the distance function over the parameters of the fundamental matrix and the 'true' positions of the correspondences. The first-order approximation of the reprojection distance, or the Sampson-error has also been applied by Torr in [20] and by Zhang in [23]. A detailed survey and performance comparison of different methods has been recently reported in [1].

The basis of most direct methods is the brightness constancy assumption, often used in the linearized form of Horn and Schunck, [8]. Direct methods as a paradigm for recovering motion was first proposed by Horn and Weldon in [9]. Rather than use point correspondence, they advocated estimating motion parameters directly from image gradient information (Irani and Anandan later generalized this to any image measurable [10]). Bergen *et al*. proposed a linear approach to estimating an affine transformation relating two views, in [2]. The general framework laid out in their paper was to become the blueprint for later methods: A coarse-to-fine approach ([3]) with iterations at each level minimizing the sum of squares errors between the first image and the transformed second image. This framework was used to propose a non-linear estimation algorithm for planar projective transformations by Szeliski in [18]. A linear version was later proposed, along with other motion models by Mann and Picard in [15]. For 3D models, Kumar *et al*. proposed a direct method for the plane+parallax mode in [13], and a multiple frame extension was later proposed by Irani *et al*. in [11]. Finally, a direct algorithm was also proposed by Stein and Shashua in [17], that utilized the *tensor brightness constraint* to estimate structure and motion directly from triplets of views. A hallmark of such direct approaches is sub-pixel accuracy and the so-called locking property in the presence of outliers. On the other hand, feature-based methods have several highly desirable qualities such as invariance to photometric and geometric transformations, and statistical modeling is easier with the feature-based abstraction of images.

The case being made here for a direct method to compute the fundamental matrix is two-tiered. First, using direct refinement of initial estimates, highly accurate estimates can be obtained between a pair of frames, particularly when motion is small. This is important in structure computation from video data, where frame-to-frame motion is often expected to be small, and highly accurate estimates of the fundamental matrices is imperative. Second, in cases where there aren't features that can be readily used or computed, or where aliasing exists, direct methods provide an accurate means of computing the fundamental matrix relating pairs of views where feature-based methods may run into detection or correspondence problems.

2. Pseudo-Warping of an Image

In [22], Wolberg defines image warping to be a spatial transformation defined by the geometric relationship between points in a source and target image. Clearly, no such image warping can be defined for the fundamental matrix since it relates a point in the source image to a line in the target image. However, in this section, we describe a function that allows us to transform a source image according to a fundamental matrix, using constraints from image information.

Definition A *pseudo-warp* of an image is a bijection, $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, of the image domain which is a function of the parameters of a geometric transformation and the image intensity information,

$$(x', y') = f(x, y|\mathbf{F}, I_1, I_2), \tag{2}$$

where (x, y) are the spatial indices on the original image, (x', y') are the transformed locations, $\mathbf{F}$ are the parameters of the geometric transformation between the image pair (I_1, I_2).

2.1. Brightness Constraint

Consider the function $I(x, y, t)$, mapping the imaged spatio-temporal space recorded by a camera to brightness values. Using the *brightness constancy constraint*, i.e. the brightness of a point on the object, as recorded by the camera, remains the same across time, we have,

$$I(x, y, t) = I(x + u, y + v, t + 1), \tag{3}$$

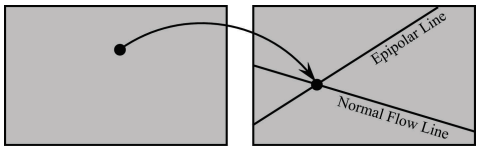

Figure 1. Pseudo-warping. Using geometric and brightness constraints in conjunction allows a one-to-one transformation of image pixels.

where u is the x-component of the optical flow, i.e. $u = x' - x$, and v is the y-component, $v = y' - y$, where $(x', y', t+1)$ is the point corresponding to (x, y, t). Expanding the right-hand side of the equation using Taylor's series and retaining only the linear terms, at the point (x, y, t) we have,

$$I(x+u, y+v, t+1) = I(x, y, t) + \frac{\partial I}{\partial x} \cdot u + \frac{\partial I}{\partial y} \cdot v + \frac{\partial I}{\partial t} \cdot 1 + \dots \tag{4}$$

Substituting Equation 3 in Equation 4, we get the well-known optical flow constraint equation of Horn and Schunk, [8],

$$I_x(x, y, t)u + I_y(x, y, t)v + I_t(x, y, t) = 0, \tag{5}$$

where I_x, I_y and I_t are the spatial (in x and y) and temporal image derivatives respectively. The optical flow constraint equation (Equation 5) can be rewritten as the equation of a line,

$$v(x, y, t) = -\frac{I_x(x, y, t)}{I_y(x, y, t)} u(x, y, t) - \frac{I_t(x, y, t)}{I_y(x, y, t)}. \tag{6}$$

This linear ambiguity is often referred to as the *aperture problem*. Since $u = x' - x$ and $v = y' - y$, we can re-write this linear constraint in terms of x' and y', the coordinates at $t + 1$ of the point corresponding to (x, y).

$$y' - y = -\frac{I_x(x, y, t)}{I_y(x, y, t)}(x' - x) - \frac{I_t(x, y, t)}{I_y(x, y, t)},$$

$$y' = -\frac{I_x(x, y, t)}{I_y(x, y, t)} x' - \left(\frac{I_t(x, y, t)}{I_y(x, y, t)} - y - \frac{I_x(x, y, t)}{I_y(x, y, t)} x \right). \tag{7}$$

Finally we can rewrite this equation in homogeneous coordinates as,

$$[x', y', 1]^T [I_x, I_y, I_t - yI_y - I_x x] = [x', y', 1]^T \cdot l'_b = 0 \tag{8}$$

where the true flow is constrained by brightness values to lie on the line l'_b.

2.2. Geometric Constraint

The fundamental matrix, $\mathbf{F}$, satisfies the constraint that for a pair of corresponding points $(x, y, 1)$ and $(x', y', 1)$,

$$[x', y', 1]\mathbf{F}[x, y, 1]^T = 0. \tag{9}$$

The epipolar line is $l'_e = \mathbf{F}[x, y, 1]^T$ and therefore for a rigid scene we have,

$$[x', y', 1]^T \cdot l'_e = 0. \tag{10}$$

Thus, given that the fundamental matrix relating two consecutive views, the correspondence at any point can then be computed,

$$\mathbf{x}' = l'_e \times l'_b, \;\; \lambda \mathbf{x}' = [x', y', 1]. \tag{11}$$

Expanding the equation we get,

$$x' = \frac{(f_4 x + f_5 y + f_6)(I_t - I_y y - I_x x) - (f_7 x + f_8 y + f_9) I_y}{(f_1 x + f_2 y + f_3) I_y - (f_4 x + f_5 y + f_6) I_x}, \tag{12}$$

$$y' = \frac{(f_7 x + f_8 y + f_9) I_x - (f_1 x + f_2 y + f_3)(I_t - I_y y - I_x x)}{(f_1 x + f_2 y + f_3) I_y - (f_4 x + f_5 y + f_6) I_x}. \tag{13}$$

Equations 12 and 13 define the relationship between corresponding points in the two images. The transformation described by these equations are analogous to other point-to-point warping such as planar affine transformations or planar projective transformations. However, unlike these purely geometric transformations, the *pseudo-warping* defined by Equations 12 and 13 differs from regular warping in two important ways. First, unlike regular warping, it requires image information for transformation, thus it will transform different images differently for the same geometric transformation (e.g. fundamental matrix parameters). Second, since the optical flow constraint equation assumes small motion, this transformation loses fidelity in warping large transformations. The *pseudo-warping* between points in the two images is shown in Figure 1.

3. Estimation of the Fundamental Matrix

In this section we describe a hierarchical scheme for estimating the fundamental matrix, while maintaining its rank deficiency. Since Equation 5 assumes small motion and since the Levenberg-Marquardt algorithm tends to get stuck at a local optima, we perform hierarchical matching using Gaussian pyramids to capture larger motions, [3]. In this scheme, sub-sampled versions of the images are used to estimate a coarse estimate of the parameters, because the same motion becomes smaller at coarser levels. These coarse estimates are then propagated to progressively finer levels of the pyramid to refine the estimate at each level of the hierarchy. Each level of the pyramid is a transformation of the image at a finer level by the 2D similarity transform $\mathbf{T} = diag(\frac{1}{2}, \frac{1}{2}, 1)$, therefore the fundamental matrix estimated at level i is propagated to a finer level $i + 1$ as $\mathbf{F}_{(i+1)} = \mathbf{T}\mathbf{F}_{(i)}\mathbf{T}$. At each level of the Gaussian pyramid, several iterations of the following minimization are performed. As in [2, 18], in order to find the optimal estimate of the fundamental matrix, we wish to minimize the

mean sum of squares of the intensity error,

$$E = \frac{1}{n}\sum_{i}^{n}[I(x_i, y_i, t) - I(x_i{}', y_i{}', t+1)]^2 = \frac{1}{n}\sum_{i}^{n} e_i^2, \tag{14}$$

over all the n corresponding pairs of pixels[1]. We use the Levenberg-Marquardt minimization (see [16] for details) to perform this non-linear minimization. We need to minimize this with respect to the parameters of the fundamental matrix, $\{f_1, f_2, \ldots f_9\}$, and as a result we need the partial differentials of e_i with respect to each of the unknown parameters. Using the chain rule, these can be computed,

$$\frac{\partial e_i}{\partial f_j} = \frac{\partial e_i}{\partial x'}\frac{\partial x'}{\partial f_j} + \frac{\partial e_i}{\partial y'}\frac{\partial y'}{\partial f_j}. \tag{15}$$

Clearly, $\frac{\partial x'}{\partial f_j}$ can be readily obtained by differentiating Equations 12, with respect to all f_j. Also, $\frac{\partial e_i}{\partial x'}$ and $\frac{\partial e_i}{\partial y'}$ can be approximated by the weighted image gradients $-e_i I_x(x_i', y_i', t+1)$ and $-e_i I_y(x_i', y_i', t+1)$. From these quantities, the approximate Hessian $\mathbf{H}$ and the gradient vector $\mathbf{G}$ can be computed,

$$\mathbf{h}_{m,n} = \sum_k \frac{\partial e_k}{\partial f_i}\frac{\partial e_k}{\partial f_j}, \quad \mathbf{g}_i = \sum_k \frac{\partial e_k}{\partial f_i}. \tag{16}$$

It is important to point out, at this stage, that all point on the lines (or curve) of Equation 8 do not exactly satisfy the brightness constancy constraint. They do not since the the Taylor series expansion used approximates the image function around a point. Intuitively, the brightness constancy assumption maps out a line in the image (normal flow line) and we know that we do not, in general, find lines of constant brightness in images. Thus different candidate fundamental matrices will give different costs according to Equation 14, which can be used to select a 'best' fundamental matrix.

3.1. Enforcing the Rank Constraint

By virtue of its construction, the fundamental matrix is known to be rank deficient. A straightforward strategy that is sometimes employed to enforce this constraint is to factorize the estimated $\mathbf{F}$ using Singular Value Decomposition, and computing the singular matrix $\mathbf{F}'$ by setting the smallest singular value to zero. This provides an estimate of the fundamental matrix that minimizes the Frobenius norm $||\mathbf{F} - \mathbf{F}'||$, [6]. However, this method is inexact and is likely to increase the error of the estimate with respect to Equation 9. To obtain the most accurate computation of the fundamental matrix, this rank 2 constraint on the fundamental matrix must be enforced *during* estimation and not 'after the fact'. For projective cameras, the fundamental matrix is rank 2, and its right and left null space is spanned by the epipoles in each image respectively, $\mathbf{e}'^T\mathbf{F} = \mathbf{F}\mathbf{e} = \mathbf{0}$. The rank constraint on $\mathbf{F}$ can be applied by factorizing $\mathbf{F}$ as the product of two 3×2 matrices, $\mathbf{A}_1$ and $\mathbf{A}_2$ as,

$$\mathbf{F} = \mathbf{A}_1\mathbf{A}_2^T. \tag{17}$$

In order to initialize the algorithm with an estimate of $\mathbf{F}$, we can use singular value decomposition to factorize the initial estimate of $\mathbf{F}^-$ into its constituent matrices $\mathbf{U}, \mathbf{D}$ and $\mathbf{V}^T$. From these, the initial estimates of $\mathbf{A}_1^-$ and $\mathbf{A}_2^-$ are

$$\mathbf{A_1^-} = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \\ u_{31} & u_{32} \end{bmatrix}, \mathbf{A_2^-} = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \\ v_{31} & v_{32} \end{bmatrix} \begin{bmatrix} d_{11} & 0 \\ 0 & d_{22} \end{bmatrix}. \tag{18}$$

It should be noted that in this representation the fundamental matrix is parameterized by 12 elements rather than the minimum possible (seven). Thus, for the Levenberg-Marquardt algorithm, we need to compute the gradient vector and approximate Hessian with respect to the 12 parameters of $(\mathbf{A}_1, \mathbf{A}_2)$, instead of the parameters of $\mathbf{F}$. This can be done in the same way as in Section 2, replacing $\mathbf{F}$ with $\mathbf{A}_1$ and $\mathbf{A}_2$ and producing (lengthier) pseudo-warping

Objective

Compute the fundamental matrix $\mathbf{F}$ given images I_1 and I_2 and an initial estimate of the fundamental matrix $\mathbf{F}^-$.

Algorithm

1. Compute $\mathbf{A}_1^-$ and $\mathbf{A}_2^-$ using Singular Value Decomposition of $\mathbf{F}^-$
2. Create a k-level Gaussian Pyramid,
3. At each pyramid level, i = 1 to k times
   - Iterate n times
     - Use $(\mathbf{A}_1^-, \mathbf{A}_2^-)_{(i)}$ as initial estimates for the Levenberg-Marquardt algorithm.
     - Use the error, gradient vector and approximate Hessian to find the updates $\Delta\mathbf{A}_1$ and $\Delta\mathbf{A}_2$ and compute $(\mathbf{A}_1^+, \mathbf{A}_2^+)_{(i)}$ iteratively.
     - Set $(\mathbf{A}_1^-, \mathbf{A}_2^-)_{(i)}$ equal to $(\mathbf{A}_1^+, \mathbf{A}_2^+)_{(i)}$, the final estimate from the minimization routine.
   - Propagate $(\mathbf{A}_1^+, \mathbf{A}_2^+)_{(i)}$ to $(\mathbf{A}_1^-, \mathbf{A}_2^-)_{(i+1)}$.
4. Compute $\mathbf{F}$ from $(\mathbf{A}_1^+, \mathbf{A}_2^+)_{(n)}$ (using Equation 17).

Figure 2. Direct Estimation of the Fundamental matrix

[1] As in other direct approaches, if a correspondence lies outside the spatial support of the image (e.g. when the epipolar line and the normal flow line are approximately parallel) the points are ignored.

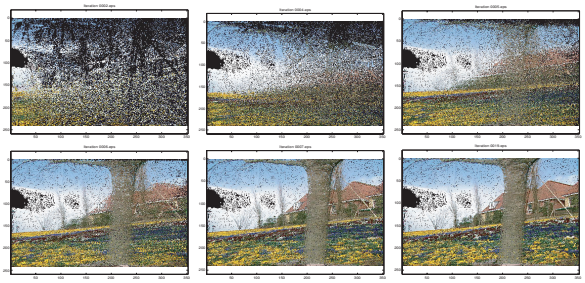

Figure 3. Pseudo-warped images during minimization. As the minimization algorithm progresses, and the estimate of the fundamental matrix improves, the pseudo-warped image begins to more closely resemble the target image (progressing from left to right, top to bottom).

functions analogous to Equations 12 and 13. To propagate $(\mathbf{A}_1,\mathbf{A}_2)_{(i)}$ to a finer level we have,

$$(\mathbf{A}_1, \mathbf{A}_2)_{(i+1)} = (\mathbf{T}\mathbf{A}_1, \mathbf{A}_2\mathbf{T})_{(i)}. \tag{19}$$

The final algorithm for hierarchical minimization is provided in Figure 2, while the pseudo-warped images during the minimization are shown in Figure 3.

An approximation that is often valid, particularly when frame-to-frame camera motion is small, is the assumption of affine cameras. If such a model is assumed then the form of the fundamental matrix becomes,

$$\mathbf{F} = \begin{bmatrix} 0 & 0 & f_1 \\ 0 & 0 & f_2 \\ f_3 & f_4 & f_5 \end{bmatrix}. \tag{20}$$

Making such an assumption reduces the Equation 12 to,

$$\begin{aligned} x' &= \frac{f_2(I_t - I_y y - I_x x) - (f_3 x + f_4 y + f_5) I_y}{f_1 I_y - f_2 I_x}, \\ y' &= \frac{(f_3 x + f_4 y + f_5) I_x - f_1 (I_t - I_y y - I_x x)}{f_1 I_y - f_2 I_x}. \end{aligned} \tag{21}$$

The assumption of an affine camera enforces the rank 2 condition on the fundamental matrix inherently, by virtue of its form, as well as reducing the number of minimization parameters to 5. Making such an approximation improves the numerical stability of estimation, and is usually useful as an initialization to the fundamental matrix for projective cameras.

4. Experimentation

In this section we report experiments that demonstrate accurate estimation of frame-to-frame motion using the proposed direct method. A MATLAB implementation, with 3 pyramid levels and 2 iterations per level, took an average of 20.32 seconds per image pair for images of size 352x240 pixels on an Intel P4 Processor, 2GHz with 1GB RAM.

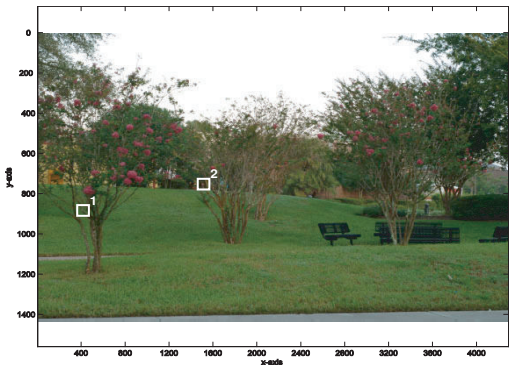

Figure 4. Sub-pixel accuracy obtained by direct method. The fundamental matrix is computed on a sub-sampled image pair and then compared at the full resolution. Zoomed in views of Area 1 and 2 are shown in Figure 5.

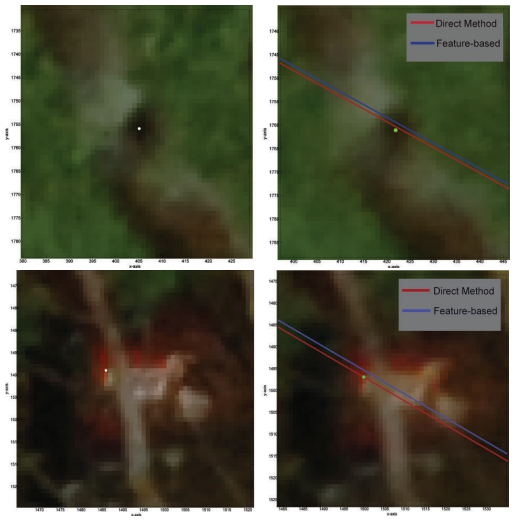

Figure 5. Zoomed in view of Area 1 and 2 in Figure 4.

We do not perform any post-processing of pseudo-warping results, although simple interpolation or in-painting algorithms can be used to obtain more visually appealing renderings. We compare the results of this algorithm against standard algorithms, such as the normalized eight-point algorithm [6] and MAPSAC [20]. To ensure fidelity of comparison, publicly available code was used for the algorithms being compared, from several libraries (Torr [21] and Kovesi [12]).

In order to show the high (sub-pixel) accuracy of the proposed approach, we captured an image pair at high resolution (2848x4288 pixels) captured by a Nikon D2x camera (as shown in Figure 4). We carefully marked corresponding points in each image pair and then resized the images to one eighth the original size (267x402 pixels). Fundamental matrices were then computed from these pairs and the residual error was computed w.r.t the propagated ground-truth correspondences. From Figure 5, it is evident that our estimate of

Table 1. Error Comparison for the Feature-based and Direct estimation of the Fundamental Matrix

Name	Direct-based SSD	Featured-based SSD
Sweater	41689.1	42349.2
Bush	12101.1	13480.1
Floor	9633.7	13746.8
Grass	8253.5	29624.4
Wall	20810.7	21016.9
Towel	21134.1	26522.6
Pot	37464.7	37852.3
Cup	27933.8	32607.4

the fundamental matrix provides higher sub-pixel accuracy compared to MAPSAC.

We also present results of estimation of frame-to-frame fundamental matrices on the flower garden sequence. We initialize our method with the MAPSAC estimated fundamental matrix. We then minimize the mean error between the pseudo-warped and the target image using two levels of pyramid with two iterations per level. Despite small frame to frame motion and accurate initial estimates from MAPSAC, the proposed approach still manages to provide refinements on the estimate. The final mean absolute error of intensity of each frame for both the methods is illustrated in Figure 6. It is evident from the figure that direct method results in a lower mean absolute error compared to MAPSAC based estimation. The direct method also results in lower accumulative error compared to MAPSAC based estimation, between consecutive frames of the flower garden sequence as shown in Figure 7.

Finally, the sum of squared difference for the different images in our test set is summarized in Table 1. Figure 8 provides an qualitative error comparison of the pseudo-warped images of the estimated fundamental matrices using the feature based and direct estimation methods. The leftmost figure in each row of the test image shows the target image, the center figure depicts the pseudo-warped image of the fundamental matrix estimated using feature based method, and the right image shows the pseudo-warped image of the fundamental matrix estimated using proposed direct method. Figures 8(a), and 8(b) are images with blur, where the direct method further refines the fundamental matrix estimated using the feature based method, hence reducing the error. While, Figure 8(d) is an image of a pot with crisp features; even then the direct method has slightly improved results compared to the feature based method.

We also compared the symmetric epipolar distance error for the fundamental matrices estimated using the feature-based and the proposed direct method in Table 2.

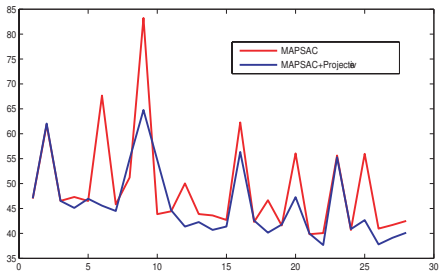

Figure 6. Flower Garden Sequence Error. The fundamental matrix between each pair of adjacent frames was computed using MAPSAC, which was then refined by the proposed approach. The two estimated matrices were then used to pseudo-warp the source frame. The mean absolute error of intensity is shown for each frame for both the MAPSAC estimate as well as the direct refinement.

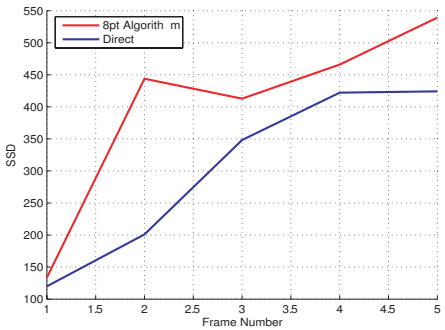

Figure 7. Accumulative error. Error between the first frame in the flower garden sequence and the next 4 frames. The error is minimized using direct refinement.

Table 2. Symmetric Epipolar Distance Error for Feature-based and Direct estimated Fundamental Matrix

Name	Direct-based Error		Feature-based Error	
	Matching	Manual	Matching	Manual
Sweater	14.54	24.67	7.82	39.14
Towel	9.42	37.73	7.74	46.53

5. Conclusion

Direct approaches have been proposed for computing many 2D parametric and quasi-parametric motion models, such as affine transformations, planar projective transformations (homographies), spline based motion, and the trifocal tensor. However, hitherto, an algorithm that computes the fundamental matrix *directly* has not been proposed. By defining a *pseudo-warping* operation with respect to an image and a fundamental matrix, we propose such as approach. On the debate of feature-based approaches vs direct approaches, we acknowledge the advantages and limitations of both methods, and adopt the view that the best results are obtained by allowing both approaches to complement each other, rather than using either in isolation.

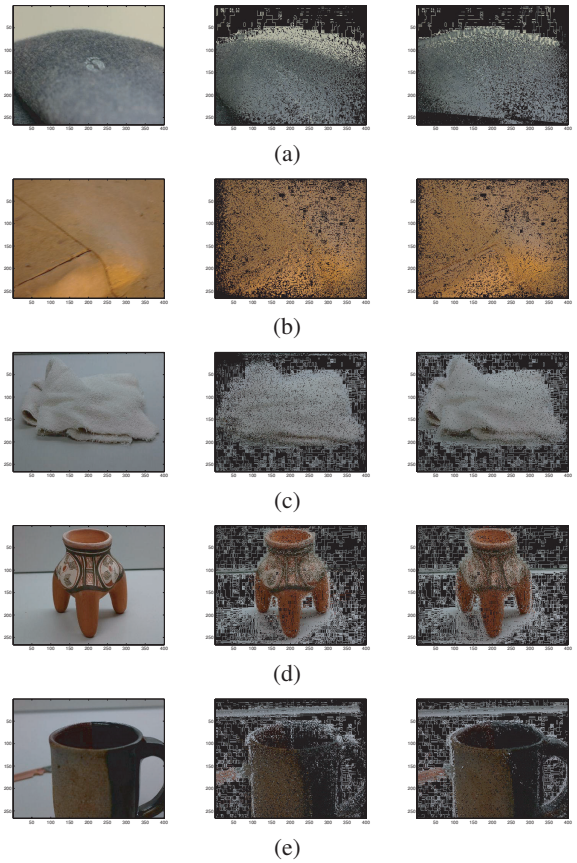

Figure 8. Comparison of the pseudo-warped images. (a) Sweater (b) Floor (c) Towel (d) Pot (e) Cup. Left to right: Target Image, MAPSAC estimation based pseudo-warped image, and Direct estimation based pseudo-warped image.

Clearly, in scenes where salient features are absent, such as blurred images, or where correspondence is difficult to estimate, such as textured images with aliasing, direct approaches offer a means to significantly improve estimates produced by feature-based approaches. On the other hand, scenes where significant transformations (both spatial and colorimetric) exist are challenging, perhaps impossible, for direct methods to reliably operate.

Acknowledgements

This research was funded by the US Government VACE program.

References

[1] X. Armangué and J. Salvi, "Overall view regarding fundamental matrix estimation," *Image and Vision Computing*, 2003.

[2] J. Bergen, P. Anandan, K. Hanna, R. Hingorani, "Hierarchical Model-Based Motion Estimation," *European Conference on Computer Vision*, 1992.

[3] P. Burt, and E. Adelson, "The Laplacian Pyramid as a Compact Image Code," *IEEE Transactions on Communications*, 1983.

[4] O. Faugeras, "What can be seen in 3D with an uncalibrated Stero rig?," *European Conference on Computer Vision*, 1992.

[5] R. Hartley, "Estimation of relative camera positions for uncalibrated cameras," *European Conference on Computer Vision*, 1992.

[6] R. Hartley, "In Defense of the Eight-Point Algorithm," *IEEE Transactions on Pattern Recognition and Machine Intelligence*,1997.

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

[8] B. Horn and B. Schunck, "Determining Optical Flow," *Aritificial Intelligence*, 1981.

[9] B. Horn and E. Weldon Jr., "Direct Methods for Recovering Motion," *International Journal of Computer Vision*, 1988.

[10] M. Irani and P. Anandan, "All About Direct Methods," *Vision Algorithms: Theory and practice*, Springer-Verlag, 1999.

[11] M. Irani, P. Anandan, and M. Cohen, "Direct recovery of planar-parallax from multiple frames," *Vision Algorithms*, 1999.

[12] P. Kovesi, "MATLAB and OCTAVE Functions for Computer Vision and Image Processing," `http://www.csse.uwa.edu.au/~pk/research/matlabfns/`, University of Western Australia, 2000.

[13] R. Kumar, P. Anandan, and K. Hanna, "Direct Recovery of Shape from Multiple Views: A Parallax Based Approach," *International Conference on Pattern Recognition*, 1994.

[14] H. Longuet-Higgins, "A Computer Algorithm for Reconstructing a Scene From Two Projections," *Nature*, 1981.

[15] S. Mann and R. Picard, "Video Orbits of the Projective Group: A Simple Approach to Featureless Estimation of Parameters," *IEEE Transactions on Image Processing*, 1997.

[16] W. Press, B. Flannery, S. Teukolsky, W. Vetterling, "Numerical Recipes in C: The Art of Scientific Computing," *Cambridge University Press*, 1988.

[17] G. Stein and A. Shashua, "Model-Based Brightness Constraints: On Direct Estimation of Structure and Motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000.

[18] R. Szeliski, "Image Mosacing for Tele-Reality Applications," *IEEE Workshop on Applications of Computer Vision*, 1994.

[19] P. Torr and A. Zisserman, "Feature Based Methods for Structure and Motion Estimation," *Workshop on Vision Algorithms*, 1999.

[20] P. Torr and A. Zisserman, "Robust computation and parameterization of multiview relations," *IEEE International Conference on Computer Vision*, 1998.

[21] P. Torr, "A Structure and Motion Toolkit in Matlab," *Technical Report MSR-TR-2002-56*, 2002.

[22] G. Wolberg, "Digital Image Warping," *IEEE Computer Society Press*, 1990.

[23] Zhengyou Zhang, "Determining the Epipolar Geometry and its Uncertainty," *International Journal of Computer Vision*, 1998.