M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 310
Appeared in Proceedings of Mustererkennung 1994, 16. Symposium der DAGM, Vienna, Austria, Tagungsband. Informatik Xpress 5, pages 186–196.

Recovery of Relative Affine Structure Using the Motion Flow Field of a Rigid Planar Patch

Nassir Navab and Steve Mann
MIT Media Laboratory,
20 Ames street,
Cambridge MA
02139, USA

Abstract

Recently there has been much interest in 3D reconstruction from uncalibrated views. *Relative Affine Structure* has been proposed [1] as an alternative framework to projective reconstruction. Proposed algorithms in [1, 2] were based on an off-line feature corresponding process. This paper extends the relative affine structure from motion to a 'correspondenceless'[1] framework.

A 'projected planar patch flow field (pflow)'[3], defined by the image motion of a rigid planar patch, is used to estimate the eight *pure parameters* [4] that describe the 2D projective transformation between the current and reference frame. The 'pflow registration' [3] brings into very precise alignment any two pictures of the same plane, even if taken by different cameras with different calibration.

However, registration will fail for objects in the scene that are not on the particular chosen plane. By exploiting this registration failure, we may determine the relative affine structure of the scene. Applying 'pflow registration' to an image sequence, one will observe that in the registered sequence objects on the plane of interest are not moving, while those that are not on the plane will move with a velocity proportional to their distance from the plane and inversely proportional to their depth in the camera coordinates. The optical flow between the registered images is an epipolar field and yields to the recovery of epipolar geometry and relative affine structure. Results on a sequence of real images helps the understanding of the process and shows that this approach is quite promising.

1 Introduction

For a quite long period photogrammeters and researchers in computer vision were concerned with the problem of reconstruction of a scene, from multiple view points. The classical *structure from motion* algorithms were suffering from need of precise camera calibration. Different off-line calibration processes have been proposed [5, 6, 7] where a calibration grid was used for recovering camera intrinsic parameters (focal length, principal point and aspect ratio). However, each time camera parameters were modified, there was a need to re-calibrate. Image sequences coming from different (probably unknown) sources could not be used for recovery of scene structure. Recently there have been some attempts at self calibration using point and line features in the scene [8, 9, 10, 11]. Unfortunately most of these algorithms are reported to be quite sensitive to noise and need good feature correspondences.

One of the first feature point reconstruction from an uncalibrated image sequence[12] was based on a parallel projection model rather than a full perspective camera model and yielded an affine structure from motion. The projective structure from motion from perspective views was later proposed in [13, 14, 15, 16]. Since then, 3D reconstruction from uncalibrated views has become an important subject of research in computer vision.

Relative Affine Structure [1] has been proposed as an alternative framework to projective reconstruction. The framework is captured by a single equation relating an invariant structure, a homography (image motion due to motion of some arbitrary rigid plane and changes in camera intrinsic parameters), and image correspondences. [17, 1] give an algebraic description of 'relative affine depth'. Using this description they show that the framework of relative affine structure is a unified framework that includes by generalization and specialization the Euclidean, projective and affine frameworks. Furthermore, in computational terms the relative affine result requires fewer corresponding points and fewer calculations than the projective framework, and is the only next general framework after projective when working with perspective views.

Algorithms proposed in [1, 2] were based on an off-line feature corresponding process. In this paper we present a Relative Affine Reconstruction method which requires no correspondences. Since finding feature correspondences is known to be a difficult problem, this work presents an important improvement to the existing algorithms. Similar approaches have been taken in parallel in [18, 19, 20]. Even though their approaches and methodologies are different these works give similar results. This shows the amount of interest and enthusiasm on this subject within the computer vision community.

1.0.1 Notation:

We consider object space to be the three-dimensional projective space $\mathcal{P}^3$, and image space to be the two-dimensional projective space $\mathcal{P}^2$. An object (or scene) is

[1] In this paper, single quotes shall denote terminology introduced by the authors here or elsewhere.

modeled by a set of points. Let $\psi \subset \mathcal{P}^2$ denote arbitrary views of the object. Given two views with projection centers $O, O' \in \mathcal{P}^3$, respectively, the epipoles are defined as the intersection of the line $\overline{OO'}$ with both image planes. Because the image plane is finite, we can assign, without loss of generality, the value 1 as the third homogeneous coordinate to every *observed* image point. When only two views ψ, ψ' are discussed, then points in ψ are denoted by p, their corresponding points in ψ' are denoted by p', and the epipoles are $v \in \psi$ and $v' \in \psi'$. The symbol $\cong$ denotes equality up to a scale, GL_n stands for the group of $n \times n$ matrices, and PGL_n is the group defined up to a scale.

1.1 Projective Planar Patch Registration

1.1.1 Video Orbits: relationship between two pictures of the same scene:

In general, the relationship between two pictures of the same scene is quite complicated. However, there are some interesting cases where this relationship is quite simple. For example, if, in a static scene, the center of projection (COP) of the camera remains fixed, then (ignoring edge effects – considering only the regions of common support in both images) the images may be related to each other by projecting one onto the unit sphere, rotating, and then projecting back onto a plane. Even if the two pictures are taken with different cameras that have different calibrations then the two images may still be related to each other by a 2-D plane projective coordinate transformation[2]. If we regard the set of images as operands, and the set of all possible 2-D plane projective coordinate transformations as operators, we have a so-called *group action*[21]. The *orbit* of a particular image is defined as the set of all possible images that can be generated by applying to it all possible 2-D plane projective coordinate transformations. Since two images taken with the fixed COP are plane-projections of one another, then they lie in the same orbit of the projective group [22].

A similar situation holds when we observe a rigid planar patch from two different cameras, even if the cameras have different calibrations. It is this simple relationship between the two pictures that we exploit, for where such a simple relationship exists, we may infer that there is a rigid planar patch present, and where such a simple relationship fails to exist, we can infer the presence of objects that do not lie on a plane. The extent to which two images do not lie in the same orbit gives us an indication as to how much they deviate from the assumed rigid planar patch. If we segment the image into small regions, we may infer from each of these the degree to which it fits a global planar surface.

1.1.2 Virtual Bellows and Virtual Graticule:

We suppose, for a moment, that our two cameras are the type of camera with full movements, wherein the lens board is connected to the film holder with flexible bellows.[3] Let us also suppose that both cameras have a ground viewing glass upon which there is etched a graticule with uniformly spaced lines.

If we place a light source behind one (call it camera #1) of the two cameras, so that the graticule is projected onto an arbitrary scene, then camera #1 "sees" a uniformly spaced graticule regardless of the scene geometry (in the same way a spotlight operator always sees a circle, regardless of the scene onto which the light is being directed). In particular, if the scene includes a rigid planar patch, then the pattern on that patch as viewed from camera #1 is a uniformly spaced graticule, regardless of the orientation of the plane with respect to camera #1. Camera #2, however, "sees" a projected graticule, where the spacing of the lines is no longer uniform, but, rather, "chirps". A "chirp" is a signal that would otherwise be periodic, but instead the frequency increases or decreases with time or space (the most familiar examples being the sound a bird makes, or the sound of a slide whistle where the frequency is changed while blowing on the whistle). In our scenario, the local spatial frequency of the image formed on the viewing glass of camera #2 is not constant, but, rather, may increase or decrease as we move across the viewing glass.

1.1.3 The pchirp and 'dechirping':

Projected periodicities on planes give rise to a local spatial frequency that rises or falls in a very particular manner. A projection of a planewave (sinusoidally varying image) is known as a 'pchirp'[23]. The process of applying the homography of the plane that makes these pchirps periodic is referred to as 'dechirping'[22], because this homography projects the chirping structure back to a harmonic structure (planewave) for which the *chirp rate* is zero. In the case of the *virtual graticule*, where the periodicity is virtual (not necessarily actually present in the image, but, rather, projected from an assumed graticule or lattice of another camera), the process of 'dechirping' may still be carried out, but between two images, rather than with a single image. Thus *registering* two images can be given the interpretation of 'dechirping' one with respect to the other. This underlying philosophy is a generalization of the use of Fourier theory for cross correlation and the underlying notion that planewaves are the eigenfunctions of linear space invariant systems.

1.1.4 Projective Planar Patch Flow (pflow):

The problem of high-resolution image reconstruction from lower-resolution image sequences [24] is one which involves *image registration*, and motion estimation. In particular, many previous models of image registration [24] used a pure translational model, wherein one image was assumed to be a shifted version of the next or previous image. In block matching [25], a similar assumption is made: a small block in one image being approximated by a translated block in the next or previous image.

More sophisticated motion models allow for rotation as well as translation [26], or even use a six-parameter 2-D affine model (which also allows for shear).

We propose an eight-parameter motion flow field. Whenever a model is constructed, there is a tradeoff between the number of parameters and the accuracy of the model. With too many parameters, a model will tend to capture "noise" and result in an underconstrained or intractable problem. However, we have found that the proposed eight parameter motion flow field nicely captures the essence of the 3-D motion of a rigid planar patch.

[2] ignoring the effects of limited depth of field (focus), diffraction, and the like

[3] Such a camera is frequently referred to as a *view camera*.

Other researchers (see for example [27]) have proposed eight-parameter flow field models such as bilinear optical flow, or "pseudo-perspective"[18]. These models only approximate the effect of perspective. Recently, however, a procedure for obtaining the "exact" generalized (eight parameter) flow field due to the motion of a rigid planar patch has been proposed [3]. The proposed method is a hierarchical recursive (iterative) procedure that uses an approximate model (such as a bilinear, biquadratic, or "pseudo-perspective" model) to bootstrap a procedure using the exact model. Within the innermost loop of the process, the inexact model is in the feedforward path and the exact model is effectively in the feedback path, as depicted in Fig 1. The algorithm is implemented by a function `est_p_m` which is a pyramidal (multiresolution) process that constructs a Gaussian (or binomial) pyramid[28]. At each level of the pyramid, `est_p_m` calls a function `est_p_r` which is a recursive plane-projective-patch parameter estimater. At all times, the parameter list (denoted $\mathbf{p}$) that is passed from one function to another, is a vector of length eight comprising the *eight pure parameters* [4]. The function `est_p_r` is the part that is depicted in Fig 1. This function calls a function `est_p` recursively. The recursive procedure is as follows:

- Call the function $\mathbf{p}$= `est_p`(I,I'). This estimates the eight pure parameters from the spatial and temporal derivatives of two frames of the image sequence, the first frame being denoted, I, and the second I', and assigns the result to parameter list $\mathbf{p}$. (Usually the parameters are estimated at each level of the pyramidal hierarchy, and then passed to the next level.)

- Resample ("warp[4]") I' according to the estimated $\mathbf{p}$ that describes the generalized motion between I and I' and call this resampled image I'_1. We would expect that I'_1 looks more like E than does I'.

- Compute the residual eight pure parameters that describe the generalized motion between I and I'_1, and call these parameters $\mathbf{p}_1 =$ `est_p`(I, I'_1).

- Apply the *law of composition* afforded by the group structure[22] of plane projective geometry to obtain $\mathbf{p}_1 \circ p$. Use this composite parameter list to resample I', and call this result I'_2, which should now look a lot more like I than either I' or I'_1.

- Repeat as many iterations as necessary to reach the desired error criterion.

A simple error criterion is the sum-of-squared difference between the reference image I, and the resampled comparison image I'_k.

The function `est_p` comprises a call to a function `est_q` followed by a call to a function `q_to_p`. The function `est_q` may be a bilinear, biquadratic, or pseudo-perspective approximation to the true motion model, and, for example

[4]This term is a bit misleading in this particular case since the mapping used here takes straight lines to straight lines – nothing gets curved or bent as might be implied by the word "warp".

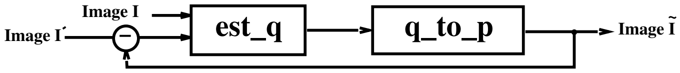

Figure 1: Plane projective projection parameter estimation process. First we estimate using an approximate model (with 8 or more parameters), such as a bilinear, biquadratic, or pseudo-perspctive model, and then we relate those parameters to the eight pure parameters.

(pseudo-perspective) may be implemented solving:

$$\sum(FF^T)\begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \\ q_5 \\ q_6 \\ q_7 \\ q_8 \end{bmatrix} = -\begin{bmatrix} \sum I_t I_x \\ \sum I_t x I_x \\ \sum I_t y I_x \\ \sum I_t I_y \\ \sum I_t x I_y \\ \sum I_t y I_y \\ \sum I_t x^2 I_x + xyI_y \\ \sum I_t xyI_x + y^2 I_y \end{bmatrix}$$

where $F = [I_x \;\; xI_x \;\; yI_y \;\; (x^2I_x + xyI_y) \;\; (xyI_x + y^2I_y) \;\; I_y \;\; xI_y \;\; yI_y]^T$ and the summations are over the entire patch (over all x and y). This equation looks a lot like the six by six matrix equation presented in Bergen [29], though we have written it in a less cumbersome form by taking the summations outside the outer product matrix with the assumption that first the outer product is taken, and then the summation acts on each of the 64 entries of the matrix producing 64 scalar quantities. The function `q_to_p` converts the parameter list returned by `est_q` into the form used by the rest of the procedure (e.g. either the *eight pure parameters* or the four corners of the bounding box of the image of the planar patch).

1.2 Relative Affine Structure

In this paper we use the results of the following three lemmas. These lemmas and their proof are discussed in detail in [1, 17]:

Lemma 1 *Let p_0, p_1, p_2, p_3 and p'_0, p'_1, p'_2, p'_3 be four corresponding points coming from four non-coplanar points in space. Let A be a collineation of P^2 determined by the equations $Ap_j \cong p'_j$, $j = 1, 2, 3$, and $Av \cong v'$. Finally let v' be scaled such that $p'_0 \cong Ap_0 + v'$. Then, for any point $P \in P^3$ projecting onto p and p', we have*

$$p' \cong Ap + kv' \tag{1}$$

The coefficient $k = k(p)$ is independent of Ψ', i.e., is invariant of the choice of the second view, and the homogeneous coordinates of M are $(x, y, 1, k)$. The scalar k is called a relative affine invariant.

Lemma 2 *vLet the coordinate transform from $P = zp$ to $P' = z'p'$ be described by $P' = M'RMP + M'T$, where R, T are the rotational and translational parameters of the relative camera displacement, and M, M' are the internal camera parameters (the transformation from camera coordinate systems to the observed coordinate systems in which $p = (x, y, 1)^\top$ and $p' = (x', y', 1)$). Given A, π, k defined in Lemma 1, let n be the unit normal to the plane π, and d_π the perpendicular distance of the origin to π, both in the first camera coordinate frame. Then,*

$$A \cong M'(R + \frac{Tn^\top}{d_\pi})M^{-1}, \tag{2}$$

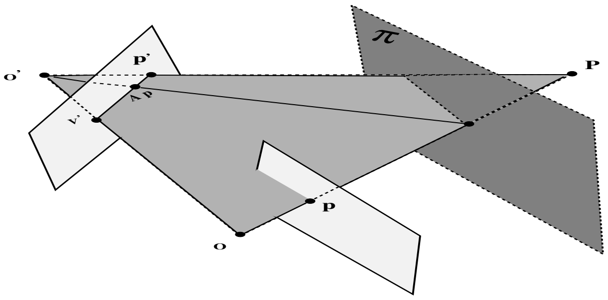

Figure 2: Image of p by a homography defined by the plane π, Ap, and image of 3D point P on the second camera, p', both lie on the epipolar plane defined by P, O and O'. Therefore p', Ap and v' are collinear.

and

$$k = \frac{z}{z_o}\alpha,$$

where z_o is the depth of $P_o \notin \pi$, and $\alpha = \alpha(p)$ is the affine structure of P in the case of parallel projection (the ratio of normal distances of P and P_o from π).

Lemma 3 *Given two views ψ and ψ', let A and v' be the homography of π and the epipoles from view ψ to view ψ', and let the views of an object point P be p and p'. Then, the least squares solution for k is given by*

$$k = \frac{(p' \times v')^T(Ap \times p')}{\| p' \times v' \|^2}. \tag{3}$$

The first point to note is that the collineation A defined in Lemma 1 is the homography due to a plane defined by 3 non collinear points in space (p_1, p_2, and p_3 in Lemma 1 are images of these three points). (Recall homography describes the 2-D image motion due to the combined effects of 3-D motion of a rigid plane and changes in camera intrinsic parameters.) In this paper, instead of using point correspondences to compute the homography, we compute homography directly by the proposed 'pflow registration' [3] described in the previous section.

The 'pflow registration' will fail for objects in the scene that are not on the particular chosen plane. To exploit this registration failure we compute the optical flow between registered images. Let us note I and I' the two images before and after motion, and let $\tilde{I}$ be the registered image, in the sense of 'pflow registration' between I and I'. For each point $\tilde{p} \in \tilde{I}$ we then have $\tilde{p} = \lambda_p Ap$. $\lambda_p = \frac{1}{A_3^T p}$ where A_3 represents the third row of the matrix A. Let δ_p be the optical flow computed between I' and $\tilde{I}$ at the point p'. Equation (1) can then be written as:

$$\begin{aligned} p' &\cong \lambda_p(p' + \delta_p) + kv' && (4) \\ &\cong \lambda_p\delta_p + kv' && (5) \end{aligned}$$

This simply shows that optical flow defines an epipolar fields and therefore by intersecting (in a least square sense) supporting lines of optical flow vectors δ_p the epipole v' can be recovered. See Fig. 2 for a simple geometrical interpretation.

Similarly (3) in Lemma 3 can then be written as:

$$\begin{aligned} k &= \frac{(p' \times v')^T(\lambda_p(p' + \delta_p) \times p')}{\| p' \times v' \|^2} && (6) \\ &= \lambda_p \frac{(p' \times v')^T(\delta_p \times p')}{\| p' \times v' \|^2} && (7) \end{aligned}$$

This shows that the relative affine depth can be obtained directly from the optical flow between the registered images $\tilde{I}$ and I'. Through some results on a pair of real images, in the next section we give a general overview of our algorithm.

1.3 Results

Figure 4 not only shows our results but also gives a summary and overview of the whole algorithm. The pair of images we used is composed of a box located on a textured tabletop. The texture has been designed in such a way that no interest points could be located with any degree of accuracy on the plane and it is almost impossible to find good feature point correspondences to compute the homography. The planar patch was tracked through a long image sequence and it was only necessary to interactively select the planar patch in one of the images. Furthermore, a very crude outline within the planar patch is all that is required. Figure 3 points out the precision of plane registration. Figure 4 includes the optical flow between registered images and the resulting epipolar field. (Once the epipolar field is recovered, erroneous optical flow caused by occlusion can be detected and removed.) The results of Relative Affine Depth recovery are shown as a relative affine depth (k) map in Fig. 4. This depth map can be used directly for many applications such as autonomous navigation, obstacle avoidance, recognition by alignment and image coding. In order to obtain an Euclidean reconstruction one has to give more apriori information. A $4x4$ projective transformation, H, relates the relative affine and transformation relates the relative affine and Euclidean structure:

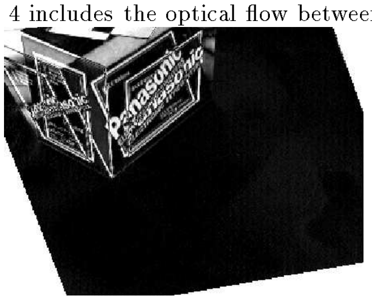

Figure 3: Absolute value difference image between reference image and resampled comparison image. To magnify small differences, the absolute difference was scaled to the interval 0..510, and clipped onto 0..255. Even after this tone scale enhancement, the error is imperceptibly small over the planar patch.

$$H \begin{pmatrix} x \\ y \\ 1 \\ k \end{pmatrix} \cong \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{8}$$

Once there is enough information, for example five 3D to 2D point correspondences, to solve (8) for H one can eas-

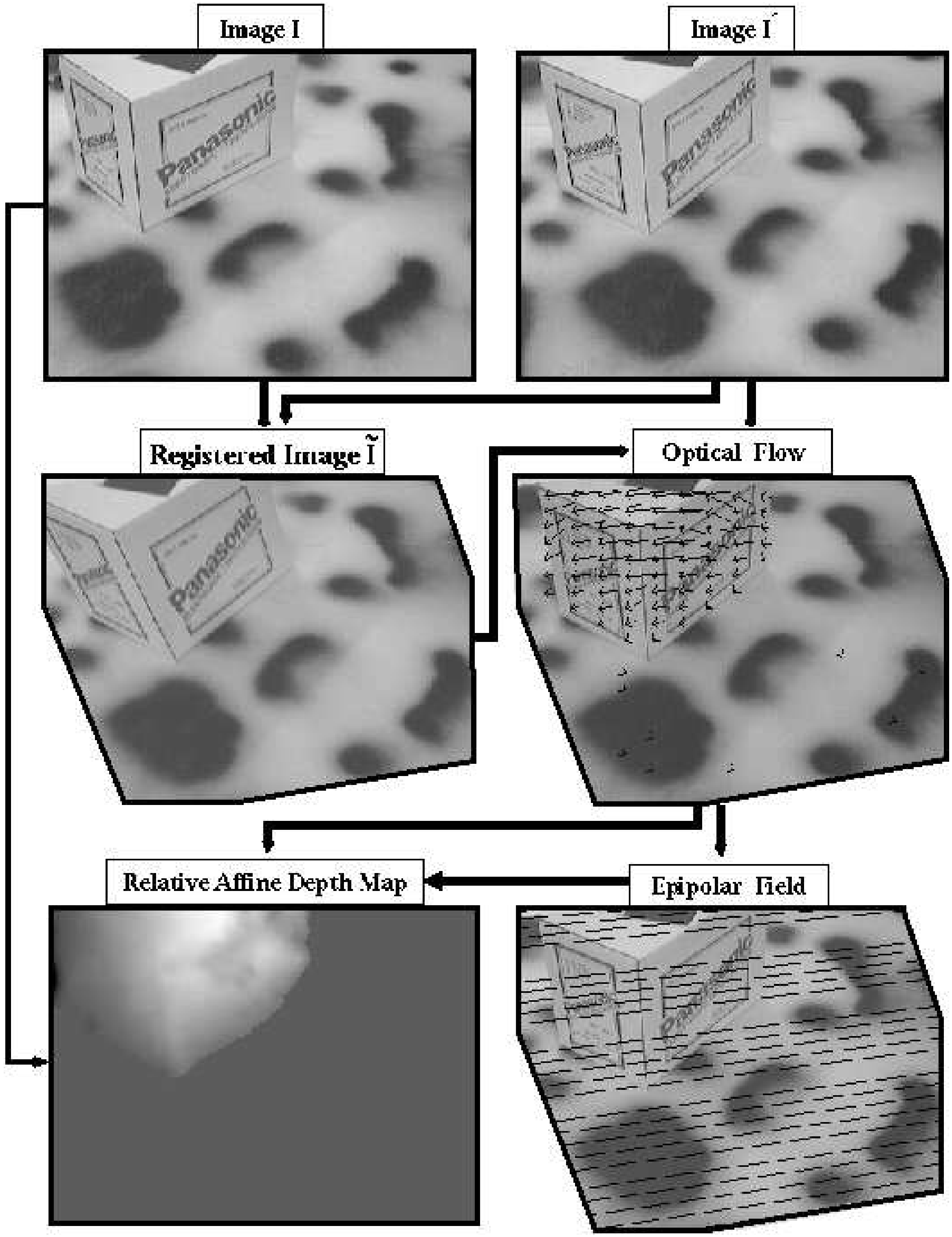

Figure 4: Overview of the algorithm: different steps of the algorithm are shown using results on a pair of real images (see text for more details).

ily recover the full 3D structure without worrying about cameras calibration or exact cameras motion.

Acknowledgments:

N. Navab was sponsored by the Nichidai Foundation fellowship, and S. Mann by B.T. and H.P. We would like to thank Aaron Bobick, Rosalind Picard, Shawn Becker and Constantine Sapuntzakis.

References

[1] A. Shashua and N. Navab. Relative affine structure: Theory and application to 3d reconstruction from perspective views. In *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, pages 483–489, Seattle, Washington, June 1994.

[2] N. Navab and A. Shashua. Algebraic derivations of relative affine structure and applications to 3d reconstruction from 2d views. In *Proc. of Commission III of the International Society for Photogrammetry and Remote Sensing ISPRS*, Munich, Germany, September 1994.

[3] S. Mann and R.W. Picard. New perspectives on the motion description of a rigid planar patch, with application to hi-res stills from video: Generalized large-displacement 8-parameter motion. TR 278, M.I.T. Media Lab Perceptual Computing Section, Cambridge, Ma, March 10 1994.

[4] R.Y. Tsai and T.S. Huang. Estimating 3-D motion parameters of a rigid planar patch, i. *IEEE Trans. ASSP*, 29(6):1147–1152, December 1981.

[5] O.D. Faugeras and G. Toscani. The calibration problem for stereo. In *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, pages 15–20, Miami, FL, June 1986. IEEE.

[6] R. Tsai. An efficient and accurate camera calibration technique for 3D machine vision. In *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, pages 364–374, Miami, FL, June 1986. IEEE.

[7] R. Lenz and R. Tsai. Techniques for calibrating of the scale factor and image center for high accuracy 3D machine vision metrology. In *Proc. Int'l Conf. Robotics Automation*, pages 68–75, Raleigh, NC, 1987.

[8] S.J. Maybank and O.D. Faugeras. A theory of self-calibration of a moving camera. *Int'l J. Comput. Vision*, 8(2):123–152, August 1992.

[9] O.D. Faugeras, Q.-T. Luong, and S.J. Maybank. Camera self-calibration: theory and experiments. In Giulio Sandini, editor, *Proc. Second European Conf. on Comput. Vision*, pages 563–578. Springer-Verlag, Lecture Notes in Computer Science 588, May 1992.

[10] R. Hartley. Projective reconstruction from line correspondences. In *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, pages 903–907, Seattle, Washington, USA, June 1994.

[11] R. Hartley. An algorithm for self calibration from several views. In *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, pages 908–912, Seattle, Washington, USA, June 1994.

[12] J.J. Koenderink and A.J. Van Doorn. Affine structure from motion. *Journal of the Optical Society of America A*, 8:377–385, 1991.

[13] O. D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig. In Giulio Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision*, pages 563–578. Springer-Verlag, Lecture Notes in Computer Science 588, May 1992.

[14] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, pages 761–764, Urbana Champaign, IL., June 1992.

[15] R. Mohr and E. Arbogast. It can be done without camera calibration. *Pattern Recognition Letters*, 12(1):39–43, 1991.

[16] A. Shashua. Projective depth: A geometric invariant for 3D reconstruction from two perspective/orthographic views and for visual recognition. *Proc. Fourth Int'l Conf. Comput. Vision*, May 1993.

[17] N. Navab and A. Shashua. Algebraic description of relative affine structure : Connections to euclidean, affine and projective structure. Technical Report 270, MIT Media Laboratory, Cambridge, USA, March 1994.

[18] M. Irani, B. Rousso, and S. Peleg. Recovery of ego-motion using image stabilization. In *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, pages 454–460, Seattle, Washington, June 1994.

[19] H.S. Sawhney. 3d geometry from planar parallax. In *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, pages 929–934, Seattle, Washington, June 1994.

[20] R. Kumar, P. Anandan, and K. Hanna. Direct recovery of shape from multiple views: a parallax based approach. In *Proc. 10th Int'l Conf. Pattern Recog.*, Israel, October 1994. IEEE.

[21] M. Artin. *Algebra.* Prentice Hall, 1991.

[22] S. Mann and R.W. Picard. Virtual bellows: constructing high-quality images from video. In *Proceedings of the IEEE first international conference on image processing*, Austin, Texas, Nov. 13-16 1994.

[23] Steve Mann and Simon Haykin. The chirplet transform: physical considerations. *To appear, IEEE Trans. on Sig. Proc.*, submitted to IEEE Trans. on Acoust. Speech Signal Process. (ASSP), 1991.

[24] A.M. Tekalp, M.K. Ozkan, and M.I. Sezan. High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration. In *Proc. of the Int. Conf. on Acoust., Speech and Sig. Proc.*, pages III–169, San Francisco, CA, Mar. 23-26, 1992. IEEE.

[25] Bernd Girod and David Kuo. Direct Estimation of Displacement Histograms . *OSA Meeting on IMAGE UNDERSTANDING AND MACHINE VISION*, June 1989.

[26] M. Irani and S. Peleg. Improving Resolution by Image Registration. *Graphical Models and Img. Proc.*, May 1991.

[27] G. Adiv. Inherent ambiguities in recovering 3D motion and structure from a noisy flow field. *IEEE Trans. PAMI*, 11(5):477–489, 1989.

[28] E. H. Adelson, E. Simoncelli, and R. Hingorani. Orthogonal pyramid transforms for image coding. In *Visual Communications and Image Processing II*, number 845 in SPIE, pages 50–58, 1987.

[29] J. Bergen, P. Burt, R. Hingorini, and S. Peleg. Computing two motions from three frames. In *Proc. Third Int'l Conf. Comput. Vision*, pages 27–32, Osaka, Japan, December 1990.