Non-Parametric Self-Calibration

David Nistér , Henrik Stewénius , Etienne Grossmann
Department of Computer Science
Center for Visualization and Virtual Environments
University of Kentucky, Lexington, USA

Abstract

In this paper we develop a theory of non-parametric self-calibration. Recently, schemes have been devised for non-parametric laboratory calibration, but not for self-calibration.

We allow an arbitrary warp to model the intrinsic mapping, with the only restriction that the camera is central and that the intrinsic mapping has a well-defined non-singular matrix derivative at a finite number of points under study.

We give a number of theoretical results, both for infinitesimal motion and finite motion, for a finite number of observations and when observing motion over a dense image, for rotation and translation.

Our main result is that through observing the flow induced by three instantaneous rotations at a finite number of points of the distorted image, we can perform projective reconstruction of those image points on the undistorted image. We present some results with synthetic and real data.

1. Introduction

Classical calibration of the intrinsic parameters of a camera is done by identifying the image points corresponding to known points on a calibration object. Self-calibration on the other hand is performed by identifying correspondences while observing an unknown static scene undergoing an unknown but rigid motion. In both classical calibration and self-calibration, the internals of the camera are assumed to follow some model with a fairly limited number of parameters. Recently, methods have been developed that perform calibration without assuming a parametric model for the camera intrinsics. This is done by calibrating each image ray separately while observing a known object undergoing known or unknown motion.

The topic of this paper is to develop a theory of self-calibration without requiring a parametric form for the intrinsics of the camera. We focus on the case of flow induced

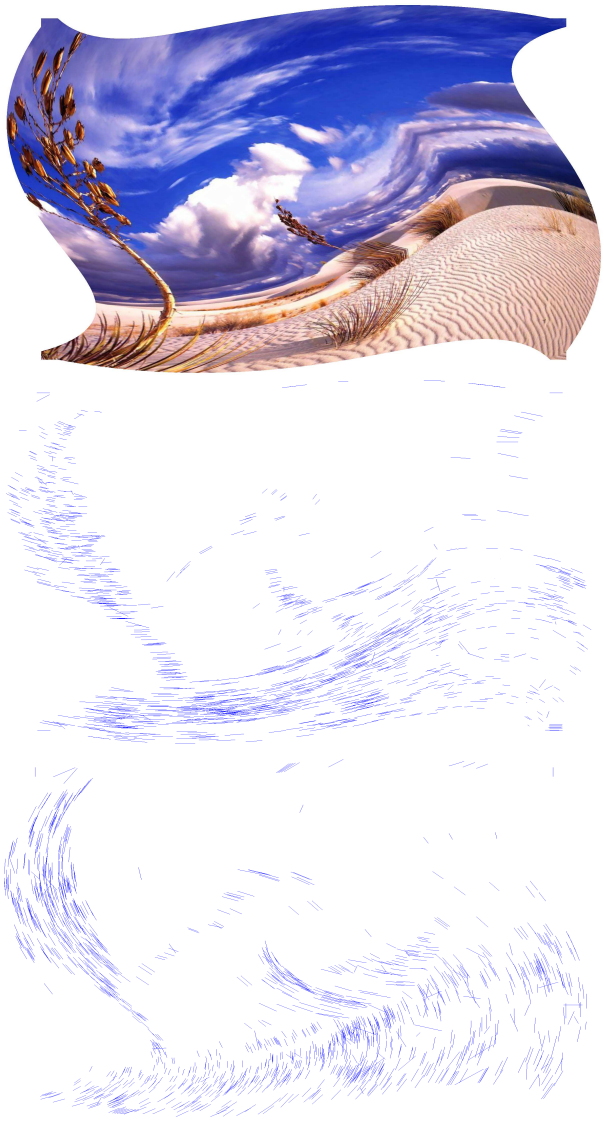

Figure 1. We develop a theory of self-calibration without assuming a parametric form for the intrinsic mapping. The task is to determine the intrinsic mapping by observing motion in the distorted image space.

by three infinitesimal rotations and observed at a finite number of locations in the distorted image. Our main result is in short that it is possible to perform projective reconstruction of the observed locations. The main result is relatively easy to state precisely, so we state it here as a preview of the theory. The following is essentially a restatement of Theorem 4:

Theorem 1 *If we observe the flow vectors $a(x)$, $b(x)$ and $c(x)$ at the points x of the distorted image, induced by three instantaneous rotations, we obtain a projective reconstruction of the undistorted image points y by setting*

$$y(x) = \begin{bmatrix} a_1(x) \\ b_1(x) \\ c_1(x) \end{bmatrix} \times \begin{bmatrix} a_2(x) \\ b_2(x) \\ c_2(x) \end{bmatrix}. \quad (1)$$

If the scene is distant enough relative to the camera geometry and motion, the camera can always be modeled as central and the motion approximated by a rotation. Some insects rotate their heads intensively, and although we do not claim that they perform self-calibration in the manner suggested by our theory, we find self-calibration with few or no assumptions on the camera an exciting prospect. In principle, one could then over time cope with any aberrations present, and in theory solve for 'everything' including an unknown scene and unknown camera.

To provide some context to our main result, we also summarize theory arising from various assumptions, including infinitesimal and finite rotations and translations, and motion observed in a finite number of locations as well as over the complete dense image sphere. With two flow-fields from a purely rotating camera, it is possible in theory to recover the intrinsics of the camera without a priori assuming a parametric form. Moreover, starting from a smooth but otherwise arbitrary intrinsic mapping, two flow-fields from a purely translating camera determine the intrinsic mapping up to a projective transformation of the image plane. Further, non-parametric self-calibration is in theory possible with two finite rotations. The case of infinitesimal translations observed at a finite number of points is equivalent to projective reconstruction in flatland. Five translations determine the points up to a 5th degree Cremona transformation, six translations up to a two-fold plus projective ambiguity and seven translations up to projective ambiguity.

The rest of the paper is organized as follows. In Section 2 we discuss related work. In Section 3 we derive our main result. In Section 4 we summarize additional theoretical results for various cases and assumptions. In Section 5 we present some results on synthetic and real images. Section 6 concludes.

2. Related Work

It is beyond the scope of this paper to give an account of the history of calibration and self-calibration. The interested reader is referred to [4, 5]. Classical calibration is performed by imaging a known calibration object [12, 14]. Self-calibration alleviates the need for a known object. It relies on observing the same a priori unknown scene in multiple images while keeping some internal parameters fixed. The best constrained case is when all intrinsic parameters are fixed [7], but it is also possible to obtain results while changing the focal length [9], or in theory even all but one intrinsic parameter [6]. Both classical calibration and self-calibration rely on a camera model known up to a small number of parameters. This constraint was dropped recently, leading to very flexible calibration using a known object undergoing known [3], or unknown [11] motion.

The goal of this paper is to investigate what can be done with a non-parametric model of a single camera when the scene is a priori unknown and not directly observable, leading to a theory of non-parametric self-calibration.

3. Main Theory

Assume that the camera produces images that are warped versions of their Euclidean counterpart given by a calibrated central perspective camera. Let the warp be described by some function $f(x)$ that indicates from which Euclidean image point y the intensity of the distorted image point x is taken, i.e.

$$y = f(x) \quad (2)$$

describes the correspondence between the distorted and the Euclidean image.

In the infinitesimal motion case, we observe a flow field $v(x)$ in the distorted image. The constraint is that when unwarped, this flow must be some underlying Euclidean flow field $m(y)$, which even for general 3D motion is more limited than a general flow. Formally, the constraint is

$$\frac{\partial f}{\partial x}(x)v(x) = m(f(x)), \quad (3)$$

where $\frac{\partial f}{\partial x}(x)$ is the matrix derivative of f at x.

We will find it convenient to think of points $y = f(x)$ in the Euclidean image as points on the unit sphere and the flow $m(y)$ as vectors based at y and tangent to the unit sphere. The flow induced at a point y in the Euclidean image by an instantaneous rotation r is then

$$m(y) = r \times y, \quad (4)$$

which when inserted into Equation (3) gives

$$Jv = r \times y, \quad (5)$$

where $J = \frac{\partial f}{\partial x}(x)$ is a 3×2 matrix.

The information about the warping function f in the neighborhood of a point x is contained in $y = f(x)$ and J. Given the observed flow vectors v we wish to recover the unwarped image locations y.

If we only observe the flow at a finite number of points, it turns out that the rotations r can only be recovered up to a common scale and multiplication by a 3×3 matrix, and that the points y can only be determined up to a projective transformation:

Theorem 2 *If we observe the flow vectors at a finite number of points x of the distorted image, induced by instantaneous rotations r, we can at best recover the undistorted points $y = f(x)$ corresponding to x up to a projective transformation, and the rotations r up to a common scale and multiplication by a non-singular 3×3 matrix describing the same projective transformation.*

Proof: To obtain the proof, we need the algebraic equality

$$[Hr]_\times H = det(H) H^{-\top} [r]_\times , \tag{6}$$

which holds for any invertible 3×3 matrix H and 3-vector r, see e.g. [5].

Let H be a 3×3 matrix describing a projective transformation and λ a non-zero scalar. Assume that y, r and J constitute a valid solution to Equation (5). Then

$$\begin{aligned} y' &= Hy/|Hy| && (7)\\ r' &= \lambda Hr && (8)\\ J' &= \lambda det(H) H^{-\top} J/|Hy| && (9) \end{aligned}$$

also constitute a valid solution, since y' is a unit vector, y' is orthogonal to the columns of J' and

$$J'v = \frac{\lambda det(H) H^{-\top} Jv}{|Hy|} = \frac{\lambda [Hr]_\times Hy}{|Hy|} = r' \times y', \tag{10}$$

by Equations (5) and (6). □

The projective ambiguity turns out to be precise:

Theorem 3 *If we observe the flow vectors at a finite number of points x of the distorted image, induced by three instantaneous linearly independent rotations r we can recover the undistorted points $y = f(x)$ corresponding to x up to a projective transformation, assuming that the warp f has a well defined non-singular matrix derivative at the points x under study.*

Proof: The constraints from Equation (5) given by three rotations can be combined into the matrix equation

$$JV = -[y]_\times R, \tag{11}$$

where V is a 2×3 matrix whose columns are the three flow-vectors v observed for a region and R is a 3×3 matrix whose columns are the three rotation vectors r. According to Theorem 2, we can choose the rotations arbitrarily to fix the projective coordinate system, without loss of generality. Hence, we choose $R = I$ and obtain

$$JV = -[y]_\times . \tag{12}$$

The matrix V would only drop rank if the flow vectors were equal up to scale, which can not happen since the rotation vectors are linearly independent. Hence, V has a nullvector that is unique up to scale. According to the assumptions, J has rank two. Therefore, the nullvector of V must also be the nullvector of JV, unique up to scale. Looking at the right hand side of Equation (12), it is seen that this nullvector must be y. Hence we can determine y uniquely as the unit length nullvector of V. Note also that once we have obtained y, Equation (12) gives us three linear equations with total rank two for each of the unknown rows of J, so that J can always be uniquely determined. □

The above proof also proves the following, which is a different way of stating Theorem 1:

Theorem 4 *Under the assumptions of Theorem 3, we can obtain a projective reconstruction of the points y by solving for each point $y = f(x)$ as the nullvector of the 2×3 matrix V whose columns are the three flow-vectors observed at the point x in the distorted image.*

4. More Theory

In this section, we summarize some additional theoretical results in order to provide some context to our main result. Due to space limitations, we will give the theorems without proof and concentrate on explaining their meaning.

Algebraic treatment of a finite number of observations has practical potential for two reasons. It does not require that the mapping is well-behaved everywhere, but only in a small number of places. Moreover, it allows for building robustness into the procedure when observing many motions, since only a finite subset of the observations from one flow need to be correct, opening up the possibility of fusing a number of projective reconstructions to achieve robustness. This is why we have chosen to concentrate on this case. However, one can make various other assumptions and obtain theory of different flavor. We list a number of cases in Table 1.

If one assumes that the motions are observed on the whole dense image sphere, and that the warping function is smooth, it only takes two finite or infinitesimal rotations to determine the intrinsic mapping, and the ambiguity is further restricted to metric:

	Finite number of observation points	Observation of complete dense image sphere
Infinitesimal Rotation	3 motions → projective	2 motions → metric
Infinitesimal Translation	5 motions → Cremona transformation 6 motions → twofold + projective 7 motions → projective	2 motions → projective
General Infinitesimal Motion	Unknown	Unknown
Finite Rotation	NA	2 motions → metric
Finite Translation	NA	At best projective
General Finite Motion	NA	Unknown

Table 1. Overview of the various theoretical cases. We focus on the top left case of flow-vectors induced by three infinitesimal rotations, observed at a finite number of points. Throughout the table, motions are assumed to be general. The (static) scene may be changed for each motion measurment. In the case of observing the complete dense image sphere, the distortion mapping is assumed to be smooth. See the text for more details.

Theorem 5 *Assume that two rotational flows of the Euclidean image sphere, with distinct axes of rotation, are observed on the whole distorted image sphere. Then the metric of the image sphere is determined uniquely.*

Theorem 6 *Assume that two finite rotations of angles β_1 and β_2 around distinct axes are observed on the whole distorted image sphere, and that neither of the ratios $\frac{\beta_i}{2\pi}$ are rational. Then the metric of the image sphere is determined uniquely.*

Note the difference in the infinitesimal motion case between the projective ambiguity with a finite number of point neighborhoods and metric reconstruction using the whole image sphere. It arises because with a finite number of observations, there are no restrictions on the flow-curves between observations. With dense observations, we can in theory trace out the flow curves. With two rotations, the intertwined systems of flow-circles then restrict the point positions to metric, see Figure 2.

Two purely translational flows also lead to projective reconstruction:

Theorem 7 *Assume that two purely translational flows on the Euclidean image sphere, with distinct directions of translation, are observed on the whole distorted image sphere. Assume that the depth map is piecewise continuous and contains only finite depths. Then the non-parametric warping function from the Euclidean image to the distorted image is determined up to a projective transformation.*

The projective ambiguity is also present in standard self-calibration with a purely translating camera, and remains regardless of how many translations are observed. The proof of the above theorem proceeds by considering the pencils of lines [10] through the foci of expansion, see Figure 3. In theory, one can use the flow to determine the lines through the foci and the structure of the two pencils, because the matrix derivative of the warp is well defined at the foci. After setting the coordinates of the foci and two additional points to define the projective coordinate system, the remaining points are then uniquely determined as the intersection of two lines from the foci.

Also the pure infinitesimal translation case can be considered with a finite number of observations and handled algebraically. Assume that we observe the effects of n pure translations in m point neighborhoods of the distorted image. As before, we assume that the distorting mapping has a non-singular matrix derivative at the points under study. We also assume that the depth values in a particular point neighborhood are different for each translation. Note that one could also study translations from the same base position, in which case the depth value in a particular neighborhood would be the same for all the motions.

The following is an equivalence result:

Theorem 8 *If we observe the flow-vectors induced by n pure infinitesimal 3D-translations at m points in the distorted image, and wish to recover the positions of the points on the image sphere, this problem is algebraically equivalent to the problem of performing projective reconstruction in flatland (1D-cameras in a 2D-world) using n points observed by m uncalibrated cameras.*

This leads to several Corollaries, which follow by applying existing theory from [1, 2, 8, 13] :

Corollary 1 *With 5 translations and a finite number of neighborhoods, the neighborhood positions are in general determined up to a 5th degree Cremona transformation.*

Corollary 2 *With 6 translations and a finite number of neighborhoods, the neighborhood positions are in general determined up to a two-fold ambiguity and a projective transformation.*

Corollary 3 *With 7 translations and a finite number of neighborhoods, the image points are in general determined up to a projective transformation.*

5. Results

We concentrate on experimental validation of Theorem 1. We investigate the accuracy with which three rotations observed at a point in the distorted image allow localization of the corresponding point in the undistorted image. More precisely, we investigate the effect of observation noise and of the configuration of the rotation axes on the localization accuracy.

An image geometry similar to that illustrated in Figure 1 was simulated by warping the flow-vectors according to Equation (5). Each point is localized according to the three flow-vectors observed at that point using Theorem 1. Although the localization is in a projective coordinate frame, we can in this synthetic experiment use the known rotations to get the correspondence between the projective and the Euclidean coordinate frame, and thereby measure the localization error as it appears in the Euclidean coordinate frame. The Euclidean x and y coordinates vary from -1 to 1 and the focal length was set to 1. The error was measured at many sensor points and the average over hundreds of samples used in the plots. When measuring noise in the observed flow-vectors, we use Decibel signal to noise ratio.

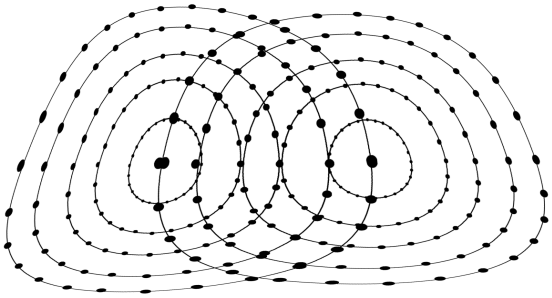

Figure 2. Illustration of Theorem 5. Intersections between flow-curves from two rotational flows places constraints on the radii of the circles corresponding to the flow-curves. This in theory allows unique recovery of the metric of the image sphere.

In Figure 4, top, the sensitivity of the point localization with respect to noise in the flow-vectors is investigated. The computation is very direct, with only a vector product constructed from the coordinates of the flow-vectors in order to reach the feature position in the projective frame. Therefore, as the figure shows, the noise in the flow-vectors is not magnified, and the noise in the output location is almost equal to the noise in the input flow-vectors.

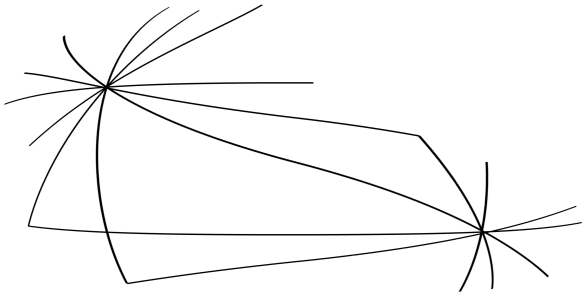

Figure 3. Illustration of Theorem 7. By observing flow in the distorted image induced by two different pure translations, non-parametric self-calibration can in theory be obtained up to a projective ambiguity, which is inherent for any number of pure translations.

We also investigate how the error is affected when the rotation axes are brought closer together. Figure 4, middle, shows how the average point localization decreases as the rotation axes separate from each other. The rotation axes were gradually varied from three coincident vectors to three vectors forming an identity matrix. In a similar fashion, Figure 4 bottom shows how the average point localization error decreases as the rotation axes move farther away from a coplanar configuration. The rotation vectors were gradually varied from $(1,0,0)$, $(0,1,0)$, $(1,1,0)$ to $(1,0,0)$, $(0,1,0)$, $(0,0,1)$. Both Figure 4 middle and bottom have two curves, one at $20dB$ and one at $40dB$. In both cases, the initial configuration (zero degrees) is degenerate, but the error decreases and reaches a minimum when the rotation axes become orthogonal.

The next experiment is illustrated in Figure 5, where results obtained with synthetically warped images are shown. The images were warped through three rotational motions. Although the images were synthetically warped a real automatic feature tracker was applied to the warped images. The tracks that survived all three motions were located in a projective coordinate system using Theorem 1. The projective frame was then set to visualize the result.

In Figure 6 results obtained with real images are shown. A consumer digital camera was rotated by hand at around one meter distance from a bookshelf. Four 1024×768 images were taken and features were automatically matched from one of the frames to the other three, resulting in three motion vectors attached to each tracked feature. The motion vectors are assumed to be flow-vectors and used in Theorem 1, which results in coordinates of each feature such that all the correctly tracked features are projective reconstructed. The projective frame was then set to visualize the

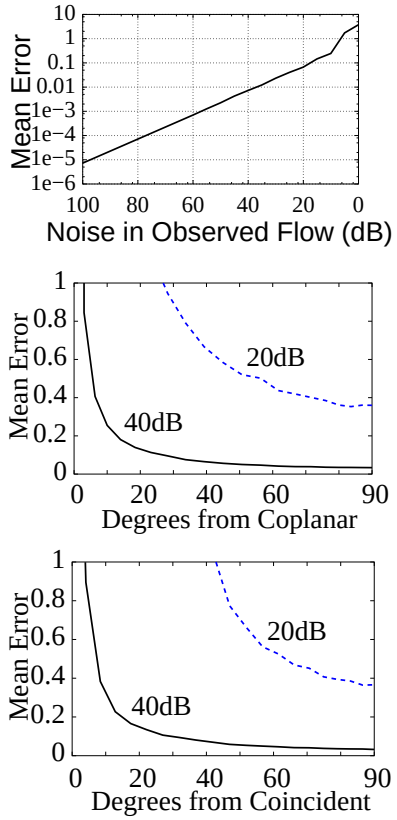

Figure 4. Top: Mean error in point localization with respect to noise in the three observed flow-vectors when the rotation axes are well separated and the rotation velocities similar. The error in the flow-vectors is given in Decibel signal to noise ratio $dB = 10\log_{10}(\frac{\text{signal variance}}{\text{noise variance}})$. The Error in point localization is measured in Euclidean image space.

result. Although some outliers cause misplaced patches in the projective reconstruction, most of the patches are correctly placed.

6. Conclusions

We have developed a theory of self-calibration without assuming a prior low-dimensional model for the intrinsics of the camera, referred to as non-parametric self-calibration. We have focused on the case of three infinitesimal rotations observed at a finite number of points. We have treated this case algebraically and shown that it is possible to recover the points at which the observations were made, precisely up to a projective ambiguity.

We have summarized a number of additional theoretical results to provide some context to our main result. Two rotational flow-fields or two finite rotations observed over the complete dense image determine the intrinsic mapping up to a metric ambiguity. Flow-fields from translation in two different directions determine the intrinsic mapping up to a projective ambiguity. Seven translations observed in a finite number of places determine the observed locations up to a projective ambiguity, while six translations lead to an additional two-fold ambiguity and five translations lead to reconstruction up to a fifth-degree Cremona transformation.

We have also presented some results with synthetic as well as real data.

References

[1] K. Åström and M. Oskarsson, Solutions and Ambiguities of the Structure and Motion Problem for 1D Retinal Vision, *Journal of Mathematical Imaging and Vision*, 12:121-135, 2000.

[2] K. Åström and F. Kahl, Ambiguous Configurations for the 1D Structure and Motion Problem, *Journal of Mathematical Imaging and Vision*, 18(2):191-203, 2003.

[3] M. Grossberg and S. Nayar, A General Imaging Model and a Method for Finding its Parameters, *IEEE International Conference on Computer Vision*, Volume 2, pp. 108-115, 2001.

[4] Gruen, A. and Huang, T. *Calibration and Orientation of Cameras in Computer Vision*, Springer Verlag, Berlin, ISBN 3-540-65283-3, 2001.

[5] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, ISBN 0-521-62304-9, 2000.

[6] A. Heyden and K. Åström, Flexible Calibration: Minimal Cases for Auto-calibration, *IEEE International Conference on Computer Vision*, pp. 350-355, 1999.

[7] S. Maybank and O. Faugeras, A Theory of Self-Calibration of a Moving Camera, *International Journal of Computer Vision*, 8(2):123-151, 1992.

[8] D. Nistér and F. Schaffalitzky, What do Four Points in Two Calibrated Images Tell Us About the Epipoles?, *Proc. European Conference on Computer Vision*, Springer Lecture Notes on Computer Science 3022:41-57, 2004.

[9] M. Pollefeys, R. Koch, L. Van Gool, Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters, *IJCV*, 32(1):7-25, 1999.

[10] J. Semple and G. Kneebone, *Algebraic projective geometry*, ISBN 0-19-8503636, Oxford University Press, 1952.

[11] P. Sturm and S. Ramalingam, A Generic Concept for Camera Calibration, *European Conference on Computer Vision*, Volume 2, pp. 1-13, 2004.

[12] R. Tsai, An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 364-374, 1986.

[13] T. Werner, Constraints on Five Points in Two Images, *IEEE Conference on Computer Vision and Pattern Recognition*, Volume 2, pp. 203-208, 2003.

[14] Z. Zhang, A Flexible New Technique for Camera Calibration, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330-1334, 2000.

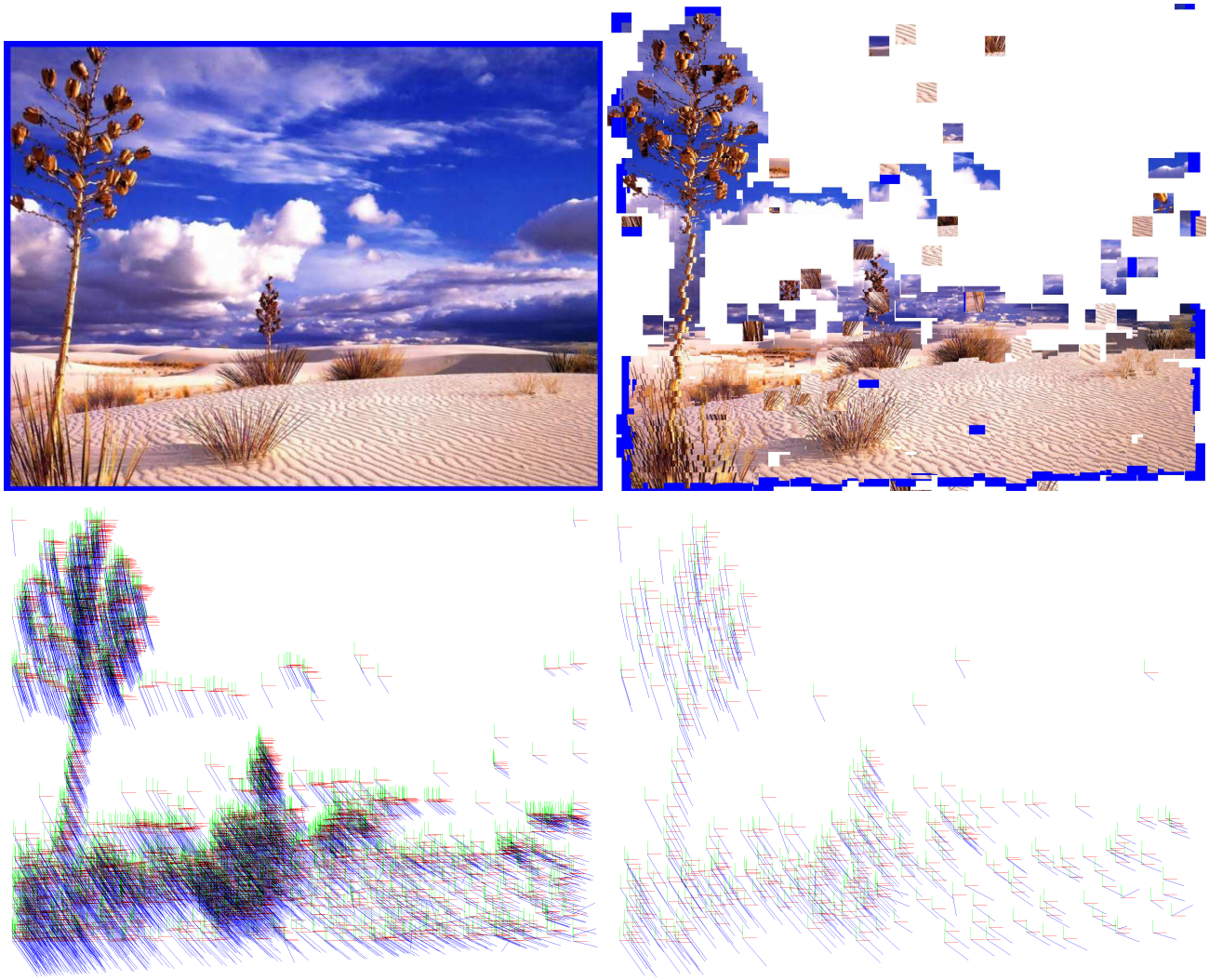

Figure 5. A result with synthetic images. The original image was warped along three rotational motions, with a sequence of ten images for each rotation. Real automatic feature tracking was applied to the warped image motions. The tracks that survived all three motions were retained. All the retained motion vectors are shown on the bottom left. To avoid clutter, a subset is shown on the bottom right. As indicated by Theorem 1, each triplet of motion vectors encodes the position of the tracked feature in a projective coordinate frame of the image. The result of placing the patch around each feature at the position indicated by the triplet of motion vectors is shown on the top right.

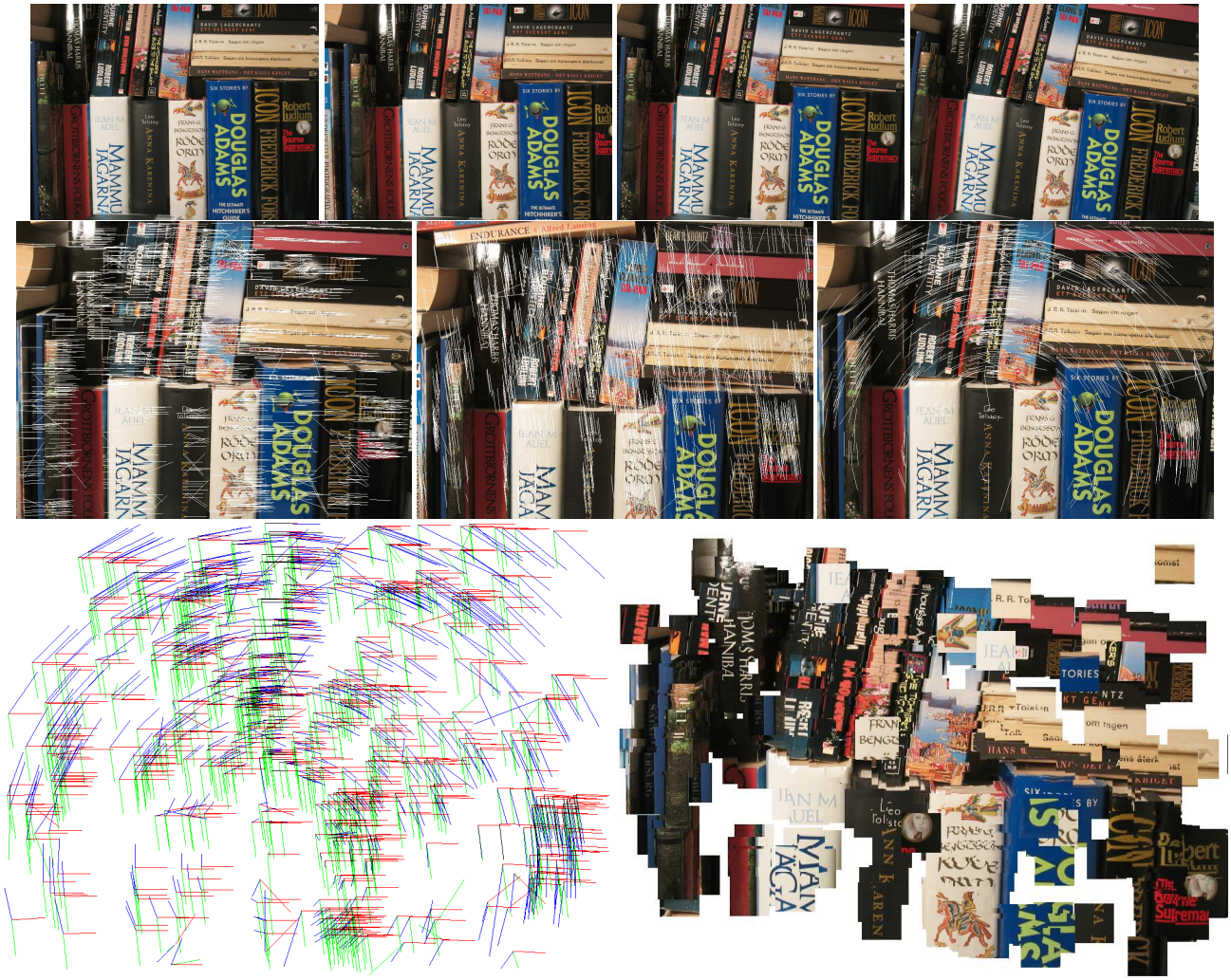

Figure 6. A result with real images. The four original images used in this experiment are shown at the top. The images were taken with a consumer digital camera. The camera was rotated with handheld rotations between the images. The correspondences, which were extracted automatically, are shown in the middle. The correspondences are centered on features. Each feature matched between all four frames gives three motion vectors. Each triplet of motion vectors, shown on the bottom left, indicates the position of the feature in a projective coordinate frame of the image. The result of placing the patch around each feature at the position indicated by the triplet of motion vectors is shown on the bottom right. The projective frame has been chosen to correspond roughly to the original images. Although some outliers cause misplaced patches in the projective reconstruction, most of the patches are correctly placed. The patches are only shown to visualize the result and not as an attempt to reconstruct the whole image.