

PRECISE REAL-TIME OUTLIER REMOVAL FROM MOTION VECTOR FIELDS FOR 3D RECONSTRUCTION

Andreas Dante and Mike Brookes

Department of E.E.E., Imperial College London, UK
andante@ic.ac.uk, mike.brookes@ic.ac.uk

ABSTRACT

Finding the correct correspondences in an image sequence is a significant task for deriving 3D structure from motion. Most research has concentrated on extracting and matching salient feature points for correspondence. Block-matching has largely been disregarded due to its significant number of correspondence-outliers and its complexity. However, nowadays real-time hardware is available to obtain block-motion vectors. We present a fast method to filter out more than 99.7% of all outliers and show that the obtained correspondences can be used to derive the 3D scene depth of real image sequences.

1. INTRODUCTION

Deriving structure from motion remains a challenging task in computer vision. It is normally tackled using a sequence of steps. The first of these steps aims to find corresponding points in a pair of images [1, 2]. In subsequent steps the Epipolar Geometry [3] encapsulated in the 3x3 Fundamental Matrix is derived from the correspondences and this matrix is decomposed into the camera extrinsic and intrinsic parameters [4]. Finally the structure is determined from the camera parameters and the correspondences [5]. The first step is very computationally intensive and therefore a problem for many real-time applications such as 3D-television.

Except for some degenerate cases [6] the above steps allow the derivation of the scene structure up to a scale ambiguity, assuming the correspondences are correct. In real image sequences the correspondences are often not correct. They are perturbed by image noise and include outliers arising from gross errors [2]. Since the calculation of the Epipolar Geometry is very sensitive to outliers, all subsequent calculations will be incorrect even with few outliers. This emphasizes the importance of rejecting outliers.

Most previous research [2, 7, 8] concentrates on extracting salient feature points from the source images as the first step in 3D reconstruction and subsequently aims to match these feature points between images. Since the features are not positioned regularly, outliers cannot be removed easily

on the basis of vector field constraints and computationally intensive (non-regular) methods must be implemented [2, 3]. In contrast to this, block matching vector fields are regular, which allows us to filter outliers efficiently using knowledge of the characteristics of realistic vector fields. Real-time hardware for block matching now exists [9, 10] and in many applications such as MPEG en-/decoders motion vectors are readily available. Thus the vector fields can be obtained at no extra cost, however they include a large proportion of outliers; these must be removed in order to obtain a correct 3D reconstruction.

The following contributions are made in this publication: (a) Description of the motion vector field in terms of the camera motion, focal lengths and object surface tilt. (b) Derivation of a reliable and fast outlier rejection scheme. (c) Presentation of a simple sub-pixel refinement method. (d) Evaluation of the new methods on synthetic and real images.

2. BLOCK MATCHING

In block matching the initial image is divided into rectangular shaped blocks. The best match for each block in the initial image is searched for in the subsequent image. The 2-dimensional search is usually limited by a window and the search steps within that window can differ depending on the precision required (e.g. full or half-pixel). The *best match* has the least absolute or squared difference of luminance values summed over all block pixels.

2.1. Sub-pixel search refinement

For a good 3D reconstruction it is important that the correspondence vectors have high accuracy. The vector precision obtained from block matching can be improved by interpolating between the position of the best match and its eight nearest neighbors (Figure 1). Using a least squares criterion we fit a quadratic surface to these nine points, of the form:

$$z(x, y) = ax^2 + bx + c + dy^2 + ey + fxy \quad (1)$$

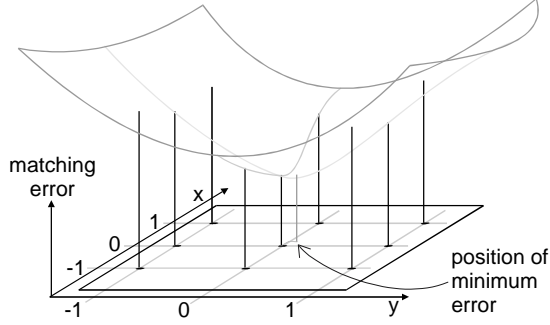


Fig. 1. Best match and eight surrounding positions in the search window as the sample points of the matching error surface

and use the minimum of the surface as the refined motion-vector.

2.2. Sources of outliers

Block matching outliers occur if the scene has little texture or large levels of noise. Outliers also occur near a center of rotation on the image plane or if the scene texture is repetitive or one-dimensional. Occlusions and depth discontinuities in the scene may also lead to outlier vectors.

3. OUTLIER REJECTION

3.1. Characteristics of vector fields

In order to filter outliers from vector fields it is important to know how vector fields are created by the camera motion and the focal length change. For a tilted planar object the vector field is described by the following equations, which are derived from the perspective projection model of rigid motion [3, 4]:

$$v_x = \frac{fk(\Upsilon R_{11} + \Lambda R_{21} + \Psi R_{31})}{\Upsilon R_{13} + \Lambda R_{23} + \Psi R_{33}} - \frac{fp_x(f - p_y m_y)}{\Theta} \quad (2)$$

$$v_y = \frac{fk(\Upsilon R_{12} + \Lambda R_{22} + \Psi R_{32})}{\Upsilon R_{13} + \Lambda R_{23} + \Psi R_{33}} - \frac{fp_y(f - p_x m_x)}{\Theta} \quad (3)$$

with

$$\begin{aligned} \Upsilon &= p_x z_0 (f - p_y \frac{dz}{dy}) - t_x (f - p_x \frac{dz}{dx}) (f - p_y \frac{dz}{dy}) \\ \Lambda &= p_y z_0 (f - p_x \frac{dz}{dx}) - t_y (f - p_x \frac{dz}{dx}) (f - p_y \frac{dz}{dy}) \\ \Psi &= z_0 \Theta - t_z (f - p_x \frac{dz}{dx}) (f - p_y \frac{dz}{dy}) \\ \Theta &= (f^2 - p_x \frac{dz}{dx} p_y \frac{dz}{dy}) \end{aligned} \quad (4)$$

where v_x and v_y are the vectors in horizontal and vertical direction, respectively, at image position $[p_x, p_y]$ and $\frac{dz}{dx}$ and $\frac{dz}{dy}$ are the surface tilts. z_0 is the surface depth at $p_x = p_y = 0$. R, t, f are the rotation, translation and initial focal length of the camera and k is the factor of change of the focal length from the first to the second view. Complex scenes may be approximated as piecewise planar segments.

Use of (2) and (3) to analyze the characteristics of vector fields leads to the conclusion: So long as the object remains in front of the camera and there are no occlusions or depth discontinuities, the motion vector fields change smoothly across the image plane independent of the camera motion. Furthermore, the relative change of a vector from a block to any of its eight direct neighbors is small, as long as the block is not near a zero vector. Such a null vector arises for example from a focal length change, a camera translation in viewing direction or a rotation around an object point.

From these observations two criteria have been developed to filter outliers, which involve only the eight blocks in the immediate vicinity of the analyzed block. This allows a parallel hardware implementation for real-time execution.

3.2. Filter criteria

The first criterion ('smooth change') is illustrated in the leftmost example of Figure 2. Here the motion vector from the central block is compared with the averages of four pairs of opposite neighbours. In this example, the two pairs shown in white give averages that are close to the central block motion vector while the shaded pairs do not. The central block is accepted as an inlier if the number of unshaded supporting pairs is above a threshold, which is determined in Section 4.

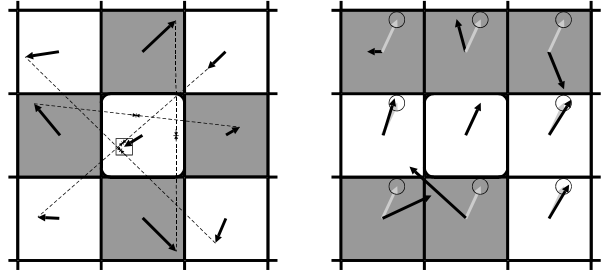


Fig. 2. Left: Smooth-change constraint; Right: Neighborhood constraint; reference block at center, supporting blocks are white

For the second criterion ('neighborhood'), illustrated in the rightmost example of Figure 2, we consider the absolute difference between the central block's motion vector and those of its eight neighbors. In this example, three of the neighbors, shown unshaded, have motion vectors that lie within a tolerance circle centered on the central vector. If the number of these supporting blocks exceeds a threshold the central block is accepted. The threshold is determined in Section 4.

3.3. Comparison of criteria

The two filter criteria are intended to remove outliers from motion vector fields of real image sequences. These how-

ever may contain depth discontinuities and a large proportion of outliers. As a result most blocks are not supported by all of their eight neighboring blocks and therefore we have to set the support thresholds of the filter criteria below the maximum.

Both criteria have different individual filter properties: The first criterion (*smooth change*) filters correctly even near a null vector, however it rejects some inlier vectors at depth discontinuities and when several of the blocks in the vicinity are outliers. The second criterion (*neighborhood*) is rather insensitive to large proportions of outliers and to depth discontinuities, since eight blocks are compared separately and not in pairs as in the first criterion. However it rejects some inlier vectors in the vicinity of a null vector.

This mutual support of both criteria implies to combine them and to reject (outlier) blocks only if both criteria fail.

4. SIMULATION RESULTS

The outlier rejection criteria have been tested and calibrated on a large set of random vector fields with injected outliers. The vector fields are created on the basis of (2) and (3) for three mutually occluding, randomly shaped planar objects under common camera motion. Each object is assigned a random depth and surface tilt. The combined scene vector field is generated by concatenating the three object vector fields. The camera position, orientation and focal lengths are changed at random such that the vector fields do not exceed the limits: $[x, y] = [-64..64, -32..32]$. This restriction models vector fields more realistically and is not a requirement for our criteria. Outliers are injected by adding a random offset vector with a minimum length of two pixels and a standard deviation of 20 pixels to a fraction of vectors.

Each of the two presented criteria has two parameters: the number of supporting blocks or block pairs and the similarity threshold. We aim to find those parameter values that give a high outlier rejection while keeping a large number of correct blocks. In order to determine the best set of parameters, several thousand tests with 100 to 10000 iterations each have been run.

We have found that a good compromise is obtained if the *smooth change* constraint (Section 3.2) is required for at least 2 pairs with a tolerance of 3% of the vector length and if the *neighborhood* constraint is alternatively required for 3 blocks with a tolerance of 8% of the vector length.

The mean result of the outlier rejection with the calibrated parameters on 10000 synthetic random vector fields is shown in Table 1.

If the proportion of outliers in the input sequence is 30% or less we obtain more than 92% of correct blocks of the input vector field while rejecting more than 99.7% of its outliers. With larger proportions of outlier blocks we obtain a proportionally higher rejection of outlier blocks but

% input outliers	10	30	50	70
% incorrectly removed blocks	1.7	7.7	24.8	58.9
% undetected outliers	0.27	0.21	0.14	0.08
% output outliers	0.03	0.10	0.19	0.43

Table 1. Mean result of our outlier rejection scheme on 10000 random scene vector fields. The last row lists the proportion of outliers in the resulting vector field after applying our method.

also a higher proportion of rejected valid blocks. In any of the measured cases (10%-70% outliers) the resulting vector field contains more than 99.5% correct vectors and less than 0.5% outliers. This makes the outlier rejection scheme very suitable for determining the Epipolar Geometry.

5. EPIPOLAR GEOMETRY

Since the computation of the Epipolar Geometry is very sensitive to outliers we reduce the number of outliers further. The RANDOM SAMPLING CONSENSUS (RANSAC) [2] is a frequently used method to remove outliers in sets of image correspondences. The number of RANSAC iterations increases exponentially with the percentage of outliers. Thus RANSAC is hard to implement in real-time for large proportions of outliers. Since we have reduced the number of outliers to less than 0.5% only very few iterations need to be run to remove the remaining outliers and to obtain the correct Epipolar Geometry.

6. ALGORITHM

The complete algorithm for deriving correspondences and the scene depths can be summarized as following:

(a) obtain motion vectors on pixel level through given hardware [9, 10]; (b) refine motion vectors to sub-pixel precision as shown in Section 2.1; (c) filter vectors using neighborhood and smoothness constraints as shown in Section 3.2; (d) run few iterations of RANSAC [2] to eliminate possibly remaining outliers. At this stage reliable correspondences are obtained. (e) Compute the Epipolar Geometry and camera parameters as described in [3, 4] and use triangulation to obtain the scene depths [5].

7. REAL IMAGE SEQUENCES

To verify the results, the algorithm has been run on ten image sequences and the Epipolar Geometry and depth for each of these has been derived. On average 1.9 outliers remained amongst the 273 correct correspondences after applying the constraints of Section 3.2. Thirty iterations of RANSAC have been sufficient to eliminate these outliers in all ten cases. The mean distance to the Epipolar lines of the more than 200 remaining features of each sequence is less than 0.25 pixels. This precision confirms the benefit of the sub-pixel refinement method described in Section 2.1. An example image of the surviving vectors, Epipolar Geometry and derived scene depth is shown in Figure 3.

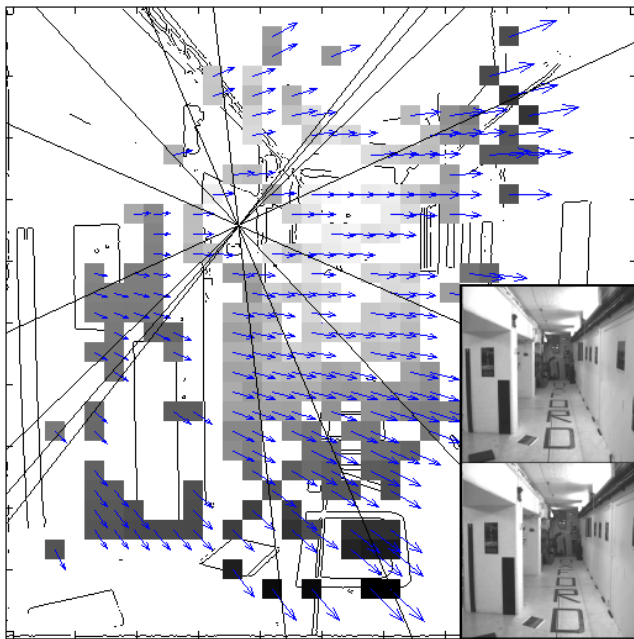


Fig. 3. Remaining correct vector field after outlier rejection, corresponding Epipolar lines, and depth map of blocks (darker=nearer); Corridor sequence [11] (original images as inset; edge image shown in the background)

It can be seen that the blocks arising from featureless walls are correctly rejected by the algorithm. As a result of the outlier rejection the Epipolar lines correctly intersect inside the image as is expected for a forward camera motion. The blocks in the center of the image are lighter (=further away) than those at the outside. This reflects the correct 3D structure of the corridor.

8. CONCLUSION

Motion vector fields have been analyzed and the equations describing the relation between camera motion, scene structure and the resulting vector fields have been derived. From

these, two criteria to filter out incorrect motion vectors have been developed. The criteria have been tested on a large set of synthetic data and it has been shown that a correspondence vector field with up to 70% outliers can be filtered such that the output vector field contains less than 0.5% outliers. Finally it has been shown that the Epipolar Geometry, and 3D depth can be derived reliably from real image sequences using our outlier rejection scheme as an initial step.

9. REFERENCES

- [1] C. Harris and M. Stephens, "A combined corner and edge detector," in *Alvey Vision Conf.*, 1988, vol. 1, pp. 147–151.
- [2] P. H. S. Torr, *Outlier Detection and Motion Segmentation*, Ph.D. thesis, Univ. of Oxford, 1995.
- [3] Z. Zhang, "Determining the epipolar geometry and its uncertainty: A review," *IJVC, Kluwer Publ.*, vol. 27, no. 2, pp. 161–195, Mar. 1998.
- [4] R.I. Hartley, "Estimation of relative camera positions for uncalibrated cameras," in *ECCV*, 1992, vol. 588, pp. 579–587.
- [5] H.C. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections," *Nature*, vol. 293, pp. 133–135, Sept. 1981.
- [6] P. Sturm, "Critical motion sequences for monocular self-calibration and uncalibrated euclidean reconstruction," in *CVPR*, 1997, vol. 1, pp. 1100–1105.
- [7] R. Koch, M. Pollefeys, and Luc Van Gool, "Realistic 3-d scene modeling from uncalibrated image sequences," in *ICIP*, 1999, vol. 2, pp. 500–504.
- [8] Peng Chang and M. Hebert, "Robust tracking and structure from motion with sample based uncertainty representation," in *Int. Conf. on Robotics and Automation*, 2002, vol. 3, pp. 3030–3037.
- [9] Yuan-Hau Yeh and Chen-Yi Lee, "Cost-effective VLSI architectures and buffer size optim. for full-search block matching algorithm," *IEEE Trans. on VLSI Systems*, vol. 7, no. 3, pp. 345–358, Sept 1999.
- [10] Seung Hyun Nam; Moon Key Lee, "High-throughput block-matching VLSI architecture with low memory bandwidth," *IEEE Trans. on Circuits and Systems II*, vol. 45, no. 4, pp. 508–512, Oct 1998.
- [11] "<http://www.robots.ox.ac.uk/~vgg/data/>," Image database of the Visual Geometry Group, Univ. of Oxford, UK.