MOTION DETECTION IN THE WITAS PROJECT

Gunnar Farnebäck, Klas Nordberg

Computer Vision Laboratory
Department of Electrical Engineering
Linköping University
SE-581 83 Linköping, Sweden

ABSTRACT

One important problem within the WITAS [1] project is detection of moving objects in aerial images. This paper presents an original method to estimate the displacement between two frames, based on multiscale local polynomial expansions of the images. When the displacement field has been computed, a plane + parallax approach is used to separate moving objects from the camera egomotion.

1. INTRODUCTION

WITAS is a research laboratory at Linköping University, currently involved in one large project focused on developing information technology for unmanned aerial vehicles (UAV's). In concrete terms, this means a small, and unmanned helicopter carrying computers, video cameras, and other electronic equipment on board, which make it capable of observing what goes on on the ground and of making decisions on the basis of these observations.

The project has two goals; to construct a particular UAV system, which is to be demonstrated before the end of year 2003, and to do high-class research on topics that are relevant for the design of such UAV's.

To reach these two goals, the project has focused on a particular operational environment, namely, roads carrying automobile traffic. The resulting system is therefore required to "understand" what happens on those roads in terms of conventional maneuvers of individual cars and other road vehicles, dangerous or otherwise exceptional maneuvers, or the structure of the traffic, e.g., congestion. It must also be able to perform tasks that are assigned by the operator or triggered by its own observations, for example to follow a certain car that flees from the scene of an apparent crime, or to assist a certain car so that it can make it through difficult traffic and get to a particular destination as quickly as possible, or to deliver a parcel to a particular point.

The authors want to acknowledge the financial support of WITAS, the Wallenberg laboratory for Information Technology and Autonomous Systems.

The UAV is supposed to perform these functions autonomously, i.e., without the direct intervention of a human operator. It is therefore not sufficient to design it for remote control of its maneuvers and of other detailed operations, the operator is only supposed to communicate general commands, often using a combination of a phrase in natural language, and pointing to a map or a video image. The most important capabilities for such a system are therefore (1) to form a model ("understanding") of scenes and events that it observes on the ground, and (2) to make prediction, planning, and autonomous decisions using that model.

The main sensor of the UAV is a camera linked to an image processing system which can analyse single images or an image sequence in order to obtain information relevant for solving a range of tasks. Typically, this includes finding and classifying individual vehicles, and measuring their velocity. Motion estimation can be used for both finding objects which are moving relative to the background, and for determining their ground velocity. Consequently, a motion estimation analysis has been implemented in the image processing system, specially designed for the restrictions imposed by the helicopter platform which carries the camera. The goal of this analysis is to allow the system to detect moving objects which later can be classified as vehicles based on other characteristics, e.g., size, and position on the ground.

This paper presents the theoretical background of the chosen motion estimation implementation. It has been made using the special purpose image processing system of the UAV system that is constructued within the project, and will in the following project phase be evaluted and tuned to the particular environment and tasks which are defined for the project.

One consequence of the camera being helicopter mounted is that it is hard to avoid vibrations, which may cause imperfect registration of subsequent frames. This makes spatiotemporal motion estimation algorithms less attractive and we have instead chosen a two-frame approach. This algorithm is based on the same ideas as an earlier disparity estimation algorithm by Farnebäck [2].

2. PRELIMINARIES

2.1. Polynomial Expansion

The first step of the signal analysis is to approximate a neighborhood of each pixel with a second degree polynomial. Thus we have the local signal model, expressed in a local coordinate system,

$$\begin{aligned} f(x,y) &\sim p(x,y) \\ &= r_1 + r_2 x + r_3 y + r_4 x^2 + r_5 y^2 + r_6 xy, \end{aligned} \tag{1}$$

or equivalently

$$f(\mathbf{x}) \sim p(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x} + c, \tag{2}$$

where

$$\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}, \ \mathbf{A} = \begin{pmatrix} r_4 & \frac{r_6}{2} \\ \frac{r_6}{2} & r_5 \end{pmatrix}, \ \mathbf{b} = \begin{pmatrix} r_2 \\ r_3 \end{pmatrix}, \ c = r_1. \tag{3}$$

The expansion coefficients $r_1, \ldots, r_6$ or $\mathbf{A}$, $\mathbf{b}$, and c are determined by a Gaussian weighted least squares fit of the signal f with the polynomial p. The details of this are out of scope for this paper but it turns out that the solution can be implemented very efficiently by a hierarchical net of 1D convolutions [3, 4].

2.2. Displacement of a Polynomial

Assume that we have an image containing an exact quadratic polynomial

$$f_1(\mathbf{x}) = p(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x} + c. \tag{4}$$

Construct a new image from the first one by a global translation $\mathbf{d}$ and expand the new polynomial

$$\begin{aligned} f_2(\mathbf{x}) &= p(\mathbf{x} - \mathbf{d}) \\ &= (\mathbf{x} - \mathbf{d})^T \mathbf{A} (\mathbf{x} - \mathbf{d}) + \mathbf{b}^T (\mathbf{x} - \mathbf{d}) + c \\ &= \mathbf{x}^T \mathbf{A} \mathbf{x} + (\mathbf{b} - 2\mathbf{A}\mathbf{d})^T \mathbf{x} + c + \mathbf{d}^T \mathbf{A} \mathbf{d} - \mathbf{b}^T \mathbf{d} \\ &= \mathbf{x}^T \tilde{\mathbf{A}} \mathbf{x} + \tilde{\mathbf{b}}^T \mathbf{x} + \tilde{c}, \end{aligned} \tag{5}$$

where the new coefficients $\tilde{\mathbf{A}}$, $\tilde{\mathbf{b}}$ and $\tilde{c}$ are given by

$$\tilde{\mathbf{A}} = \mathbf{A}, \tag{6}$$

$$\tilde{\mathbf{b}} = \mathbf{b} - 2\mathbf{A}\mathbf{d}, \tag{7}$$

$$\tilde{c} = c + \mathbf{d}^T \mathbf{A} \mathbf{d} - \mathbf{b}^T \mathbf{d}. \tag{8}$$

The key observation is that by equation (7) we can formally solve for the translation $\mathbf{d}$ as[1]

$$\mathbf{d} = -\frac{1}{2} \mathbf{A}^{-1} (\tilde{\mathbf{b}} - \mathbf{b}). \tag{9}$$

[1] Whenever something is written on the form $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$, it should be interpreted as $\mathbf{x}$ being the solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$.

3. DISPLACEMENT ESTIMATION

3.1. First Attempt

To make practical use of the observations above, we replace the global polynomial in equation (4) with local polynomial approximations. Thus we start by doing a polynomial expansion of both images, giving us expansion coefficients $\mathbf{A}_1(x,y)$, $\mathbf{b}_1(x,y)$, and $c_1(x,y)$ for the first image and $\mathbf{A}_2(x,y)$, $\mathbf{b}_2(x,y)$, and $c_2(x,y)$ for the second image. Ideally this should give $\mathbf{A}_1 = \mathbf{A}_2$ according to equation (6) but in practice we have to settle for the approximation

$$\mathbf{A}(x,y) = \frac{\mathbf{A}_1(x,y) + \mathbf{A}_2(x,y)}{2}. \tag{10}$$

We also introduce

$$\Delta\mathbf{b}(x,y) = -\frac{1}{2}(\mathbf{b}_2(x,y) - \mathbf{b}_1(x,y)). \tag{11}$$

to obtain the primary constraint

$$\mathbf{A}(x,y)\mathbf{d}(x,y) = \Delta\mathbf{b}(x,y), \tag{12}$$

where $\mathbf{d}(x,y)$ indicates that we have also replaced the global displacement in equation (5) with a spatially varying displacement field.

Simply solving equation (12) pointwise will not give very good estimates though, so in order to improve these we make the assumption that the displacement field is only slowly varying. Thus we try to find $\mathbf{d}(x,y)$ satisfying (12) as well as possible over a neighborhood I of (x,y), or more formally minimizing

$$\begin{aligned} \sum_{\{\Delta x, \Delta y\} \in I} w(\Delta x, \Delta y) \| \mathbf{A}(x + \Delta x, y + \Delta y)\mathbf{d}(x,y) - \\ \Delta\mathbf{b}(x + \Delta x, y + \Delta y) \|^2, \end{aligned} \tag{13}$$

where we let $w(\Delta x, \Delta y)$ be a Gaussian weight function. The minimum is obtained for

$$\mathbf{d}(x,y) = \left(\sum w \mathbf{A}^T \mathbf{A} \right)^{-1} \sum w \mathbf{A}^T \Delta\mathbf{b}, \tag{14}$$

where we have dropped some indexing to make the expression more readable. The minimum value is given by

$$e(x,y) = \sum w \Delta\mathbf{b}^T \Delta\mathbf{b} - \mathbf{d}(x,y)^T \sum w \mathbf{A}^T \Delta\mathbf{b}. \tag{15}$$

In practical terms this means that we compute $\mathbf{A}^T\mathbf{A}$, $\mathbf{A}^T\Delta\mathbf{b}$, and $\Delta\mathbf{b}^T\Delta\mathbf{b}$ pointwise and average these with w before we solve for the displacement. The minimum value $e(x,y)$ can be used as a reversed confidence value, with small numbers indicating high confidence. The solution given by (14) exists and is unique unless the whole neighborhood is exposed to the aperture problem.

3.2. Improved Estimation

A principal problem with the method above is that we assume that the local polynomials at the same coordinates in the two polynomials are identical except for a displacement. Since the polynomial expansions are local models these will vary spatially, introducing errors in the constraints (12). For small displacements this is not too serious, but with larger displacements the problem increases. Fortunately we are not restricted to comparing two polynomials at the same coordinate. If we have a priori knowledge about the displacement field, we can compare the polynomial at (x, y) in the first image to the polynomial at $(x+\tilde{d}_x(x, y), y+\tilde{d}_y(x, y))$, where $\tilde{d}(x, y)$ is the initial displacement field rounded to integer values.

This observation is included in the algorithm by changing equations (10) and (11) to

$$\mathbf{A}(x, y) = \frac{\mathbf{A}_1(x, y) + \mathbf{A}_2(\tilde{x}, \tilde{y})}{2}, \tag{16}$$

$$\mathbf{\Delta b}(x, y) = -\frac{1}{2}(\mathbf{b}_2(\tilde{x}, \tilde{y}) - \mathbf{b}_1(x, y)) + \mathbf{A}(x, y)\tilde{d}(x, y) \tag{17}$$

where

$$\tilde{x} = x + \tilde{d}_x(x, y), \tag{18}$$

$$\tilde{y} = y + \tilde{d}_y(x, y). \tag{19}$$

3.3. Multiscale Estimation

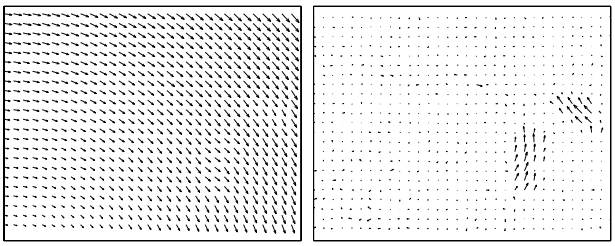
With the modified algorithm we can improve the estimates by iterating, using the estimated displacement field in one step as input to the next step. This is useful under the assumption that the input displacements in the first step have small enough errors that the new estimates are indeed improvements. One way to improve the chances for this is to iterate the algorithm over a scale pyramid. Since the estimation is most reliable for small displacements we start at the coarsest scale. The estimated field is upsampled and used as input displacements for the second coarsest scale and so on.

Figure 1 shows two frames from a test flight at Revinge. Both cars are moving slowly through the crossing while the background undergoes a substantial rotation between the two frames. The displacement field computed through iteration over three scales is shown in figure 2(a).

4. MOTION DETECTION

The final purpose of the algorithm is to detect moving objects, in particular vehicles. We cannot do this directly from the estimated displacement fields, since these include camera egomotion. To solve the problem we use the plane + parallax approach [5, 6, 7]. The idea is that the background can be approximated by a reference plane, the displacement field of which can be fit to a parametric model. After subtracting this we obtain a residual parallax displacement field where moving objects turn up and can be identified. Unfortunately also structures not lying in the reference plane cause a residual displacement, so further processing is required to distinguish these. In principle it is possible to use the fact that the parallax induced by stationary objects constitutes an epipolar field [8] but it is probably more robust and efficient to sort out potential moving objects by using other cues such as size or temporal coherence.

Fig. 1. Two frames from a test flight at Revinge.



(a) (b)

Fig. 2. Estimated displacement field (a) and residual displacement (b), subsampled and magnified.

The motion model used here is the eight parameter model,

$$\begin{aligned} v_x(x, y) &= a_1 + a_2 x + a_3 y + a_7 x^2 + a_8 xy, \\ v_y(x, y) &= a_4 + a_5 x + a_6 y + a_7 xy + a_8 y^2. \end{aligned} \tag{20}$$

The parameters are estimated by solving the weighted least squares problem

$$\arg \min_{a_1, \cdots, a_8} \sum_{x,y} w(x, y) \|\mathbf{d}(x, y) - \mathbf{v}(x, y)\|^2, \tag{21}$$

where the summation is over all points and the weights $w(x, y)$ are computed from $e(x, y)$, equation (15), as

$$w(x, y) = \frac{k}{k + e(x, y)}, \tag{22}$$

with k a design parameter. To solve (21) we rewrite (20) as

$$\mathbf{v}(x,y) = \mathbf{S}(x,y)\mathbf{p}, \quad \text{where} \tag{23}$$

$$\mathbf{S}(x,y) = \begin{pmatrix} 1 & x & y & 0 & 0 & 0 & x^2 & xy \\ 0 & 0 & 0 & 1 & x & y & xy & y^2 \end{pmatrix}, \tag{24}$$

$$\mathbf{p} = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 \end{pmatrix}^T. \tag{25}$$

Now the solution to (21) is given by

$$\mathbf{p} = \left(\sum w\mathbf{S}^T\mathbf{S} \right)^{-1} \sum w\mathbf{S}^T\mathbf{d}, \tag{26}$$

where we once more have dropped some indexing to improve the readability. The practical solution of the problem involves accumulating the coefficients of the 8×8 equation system (26) over all points and solving for the parameters.

The residual displacement field for the two frames in figure 1 is shown in figure 2(b). The residuals corresponding to the two cars are enlarged due to the averaging in equation (14).

5. FUTURE IMPROVEMENTS

Instead of fitting the eight parameter motion model in section 4 to the previously estimated displacements we can apply the primary constraint (12) (with $\mathbf{A}$ and $\mathbf{\Delta b}$ from (16) and (17)) directly to the motion model $\mathbf{d}(x,y) = \mathbf{S}(x,y)\mathbf{p}$. This gives us the least squares problem

$$\arg\min_{\mathbf{p}} \sum_{x,y} \|\mathbf{A}(x,y)\mathbf{S}(x,y)\mathbf{p} - \mathbf{\Delta b}(x,y)\|^2, \tag{27}$$

where the sum is over all points, and the solution

$$\mathbf{p} = \left(\sum \mathbf{S}^T\mathbf{A}^T\mathbf{A}\mathbf{S} \right)^{-1} \sum \mathbf{S}^T\mathbf{A}^T\mathbf{\Delta b}. \tag{28}$$

This has not been implemented yet. Since we still need to compute local displacements in order to obtain the residual parallax it is not obvious that this method is worth the extra complexity in the implementation. However, if we only need the residual field in limited regions of interest, this gives an efficient method to compute the egomotion from all points, since we avoid the relatively expensive averaging in equation (14).

6. CONCLUSIONS

We have presented a new method to estimate displacements between two frames, which combined with a plane + parallax approach can be used to detect moving objects in aerial images. Initial results look promising but work remains to optimize the implementation for the target platform and to evaluate and tune the algorithm for a wide range of environments.

7. REFERENCES

[1] WITAS web page `http://www.ida.liu.se/ext/witas/`.

[2] G. Farnebäck, "Disparity Estimation from Local Polynomial Expansion," in *Proceedings of the SSAB Symposium on Image Analysis*, Norrköping, March 2001, SSAB, pp. 77–80.

[3] G. Farnebäck, "Spatial Domain Methods for Orientation and Velocity Estimation," Lic. Thesis LiU-Tek-Lic-1999:13, Dept. EE, Linköping University, SE-581 83 Linköping, Sweden, March 1999, Thesis No. 755, ISBN 91-7219-441-3.

[4] B. Johansson, "Multiscale Curvature Detection in Computer Vision," Lic. Thesis LiU-Tek-Lic-2001:14, Dept. EE, Linköping University, SE-581 83 Linköping, Sweden, March 2001, Thesis No. 877, ISBN 91-7219-999-7.

[5] R. Kumar, P. Anandan, and K. Hanna, "Direct recovery of shape from multiple views: a parallax based approach," in *Proceedings of 12th ICPR*, October 1994, pp. 685–688.

[6] H. S. Sawhney, "3d geometry from planar parallax," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 1994, pp. 929–934.

[7] A. Shashua and N. Navab, "Relative affine structure: Theory and application to 3d reconstruction from perspective views," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 1994, pp. 483–489.

[8] M. Irani and P. Anandan, "A unified approach to moving object detection in 2d and 3d scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 6, pp. 577–589, June 1998.