Optic flow methods applied to unmanned air vehicles

Terry Cornall, Greg Egan.
Department of. Electrical. and Computer Systems Engineering
Monash University, Clayton 3800
Victoria, Australia
terry.cornall@eng.monash.edu.au

Abstract

The paper discusses aspects of the potential usefulness and limitations of methods that could use the information in the optic flow of video that can be captured and processed aboard an unmanned air vehicle. Some preliminary theoretical and experimental results from analysis of video that has been recorded by the author using video telemetry equipment designed and constructed by the author and members of the UAV group at ECSE will be presented.

1. Introduction

The hypothesis behind this work is that a small unmanned air vehicle (UAV) can maintain stable flight using only the information that can be extracted from an onboard video camera passively receiving image data from a view of the ground, using currently available computer resources and image processing methods.

This raises the corollary questions:- What information can and cannot be extracted from the video with current methods? What methods are applicable, given the constraints?

The answer to the question; "Can flight stability be achieved using only the view from the air vehicle?" is evidently "yes". It is exemplified by the obvious biological example of a human pilot flying an un-instrumented small aircraft. However the hypothesis for this work includes additional constraints, which is that this goal be achievable using a lightweight computer onboard the vehicle with no input from a human operator. The literature regarding the previous experience of others with biologically inspired tasks involving vision leads the researcher to be very wary of believing that just because a human exemplar exists, then this must mean that a computer can achieve the same goal. It is not known with complete certainty just how a human pilot achieves the goal of stable flight using only visual cues, despite many decades of work on related questions in the field of autonomous robotics and computer vision, thus it is not known how to translate the human experience into software that a computer can execute. Nonetheless, the researcher believes that there is currently enough available information about aerodynamic and computational systems and methods of computer vision to enable the hypothesis to at least be put and tested, both in a theoretical and an experimental framework.

The main focus is to achieve a sufficient mathematical and practical understanding of the theoretical aspects of the problem to be able to prove or disprove the hypothesis, and if it proves to be possible, to build and test a prototype system that embodies at least one solution to the problem.

This particular paper discusses the equipment used to capture the video from a UAV. The author then discusses how to derive an optic flow field from the onboard video, some theoretical aspects of what information is and is not available and how it can be extracted are presented and finally some examples of how such vector fields can be interpreted, using offline processing on the captured video.

2. The video and sensor telemetry system

The author has designed and built, with the considerable help of his supervisor Greg Egan, expert radio control and model aircraft designer and builder Ray Cooper and other members of the UAV group at the Electrical and Computer Systems Engineering department, a system to transmit video and GPS tracking information from the UAV to a ground receiver where it is recorded . This has allowed the author to experiment with real video sequences offline, to investigate the issues involved in designing a system that could work onboard the UAV.

The airborne part of the system consists of a video camera, a low-powered analog video transmitter, a GPS receiver and a 1200 Baud audio phase shift keyed (AFSK) modulator. The data from the GPS receiver is transmitted in the audio channel of the video transmitter, after being converted to sound by the modulator.

The ground station consists of the audio/video receiver, a digital video recorder, an AFSK demodulator and a laptop computer running software to display the GPS positioning information on a map of the area.

Various versions of the software have been written, some of which also have made use of a USB video digitization devise to display the video at the same time

as the tracking and altitude information. The video recorder records the video and the AFSK audio to allow the flight to be replayed, including the real-time tracking information, and processed offline. The author intends to apply real-time video processing techniques to this video at the ground station as a further intermediate step towards migrating the processing to onboard the UAV. At the time of this paper being produced, the offline processing is slower than real-time, although the position and altitude tracking is almost real-time with up to a one second latency due to the update time of the GPS receiver. See Fig. 1 for an example of the tracking display.
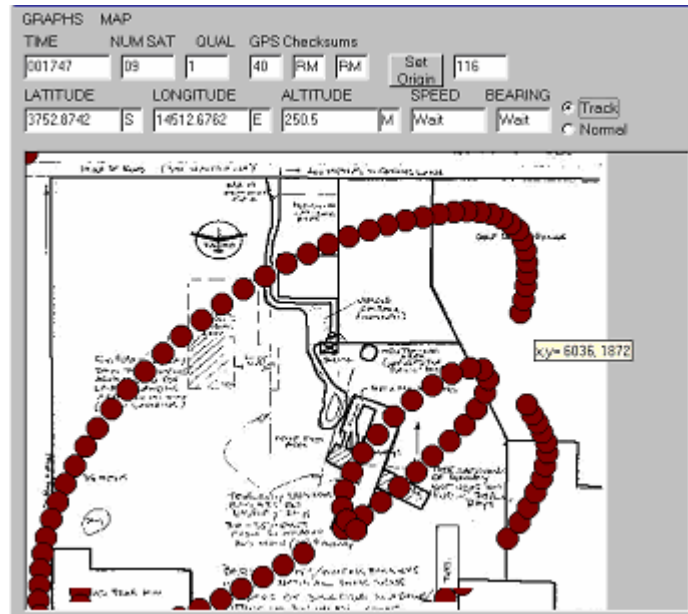

Fig. 1 GPS tracking of UAV

3. Optic flow calculations

Before discussing the theoretical aspects of the information in an optic flow field, it is appropriate to quickly overview the method used in calculating the optic flow.

The basic premises that the calculation is based upon is that the image intensity values are constant in time, and that the image intensities are continuous and differentiable. These are called the Brightness Constant Constraint, (BCC) and the smoothness constraint. As a consequence it is assumed that any changes in intensity at a point (x,y) in the image coordinates is due to the motion of the camera, or of the objects in the camera's view. [4]

It is then possible to write an equation:

$$\begin{bmatrix} \sum I_{xy}^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{bmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = -\begin{pmatrix} \sum I_t I_x \\ \sum I_t I_y \end{pmatrix}$$

relating the optic flow vector (u,v), the image intensity gradient (Ix,Iy) and the temporal changes It, in the intensity from one frame to another. [4] This equation is applied over a patch of the image and its following frame, which is the area of the summation. In the author's work to date these patches are 16x16, 32x32 and 64x64 sized blocks, non-overlapping, over the full extent of the image.

It is important to note that the equation results in an approximation to the true optic flow, and that the approximation breaks down in areas of the image where the BCC and smoothness assumptions do not hold. Also, as a consequence of the use of gradient information patches with no intensity gradient do not provide reliable results and the calculated optic flow can only provide information about the true optic flow component that lies in the direction of the gradient. The author uses a single pass algorithm at present, but there are various methods devised to improve this approximation. [2][4] However, they are iterative and add significant processing. The software that calculates the optic flow was written by the author to run under the Matlab package. Performance is adequate for offline processing, taking about 2 seconds per frame. From experience with other, non-Matlab, software that performs a similar scope of calculations the author feels that at least the optic flow calculations can be calculated in real-time. It is also expected that post-processing and information extraction from the optic flow will add significant processing time.

4. Optic flow from projected image

The process of projecting an image of the viewed objects onto the sensor plane of a camera has considerable impact on what information is and is not available from the image. The following discussion is based on the assumption of a simple perspective projection model, with a normalised focal length of 1. A good reference for the derivation and explanation of these equations is [1]. See Fig. 2. The following equations then describe the image velocities u and v in the X and Y directions, for a translation velocity V=(Vx,Vy,Vz) and a rotation O of the camera on the aircraft, looking at an object P at P= (X,Y,Z). Values denoted by an x,y,z subscript such as Vx are components in the appropriate axis of the subscripted vector. The coordinates x,y are the image coordinates of the point p=(x,y) in the image that corresponds to the point P on the object being viewed. The origin (0,0,0) of this coordinate system is the camera lens. (In reality the image plane is behind the lens, not in front of it, and the image is inverted. However, the geometry is the same if we simplify things by thinking of the image plane in front of the lens.) The Z axis is given by the direction that the camera is pointing. The image plane is in the XY plane.
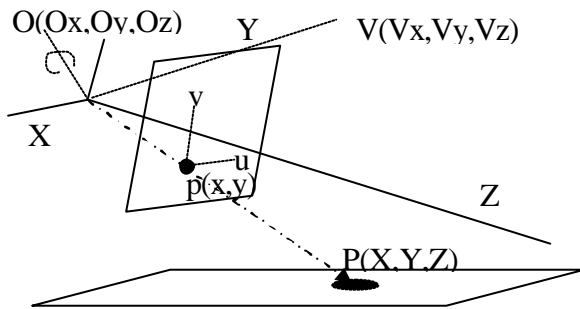

Fig. 2 Projection model.

$$(x,y)=(X/Z,Y/Z) \qquad \text{Eqn. 1}$$

$$u=(xVz/Z-Vx/Z) + (xyOx-(1+x^2)Oy-yOz) = tx+ rx \quad \text{Eqn. 2}$$

$$v=(yVz/Z-Vy/Z) + ((1+y^2)Ox-xyOy-xOz) = ty+ ry \quad \text{Eqn. 3}$$

From Eqn. 2 and Eqn. 3 we can consider the optic flow as having a translational component t and a rotational component r. Inspection of these equations reveal that if the observer's translation is zero, i.e. Vx and Vy and Vz are zero, then the values of u and v are given by rx and ry which are not dependent on the depth Z, i.e there is no depth information in the image flow for non-translational motion.

Conversely, the optic flow for a rotation does not depend on altitude, whereas the optic flow for translation does.

A point that initially caused the author some concern was whether or not a yawing rotation could be discerned from a sideways drift caused by a headwind, for example.

Consider if the rotation O is such that the component Oy is the only one non-zero, i.e. a rotation about the Y axis corresponding to a yaw then

$$u= rx = -(1+x^2)Oy$$
$$v= ry = -xyOy$$

Compare this to the case when the camera undergoes pure translation in the X direction due to Vx being a non-zero constant:

$$u= tx = -Vx/Z$$
$$v= ty = 0$$

The dependence in the rotational case of v on (x,y) and Oy compared to the zero value for v in the translational case indicates that translation due to sideways motion should be discernable from a rotation in the same plane even if there is no depth variation in the objects in the view, as could easily be the case. This does not cover the case where both Vx and Vy are non-zero, nor does it take into account that Z varies across the view of the ground, but at least at first glance the anticipated problem is not apparent. In fact , because of the relatively large value of Z at a typical altitude, the translational contributions due to all but the fastest motions, such as a dive or landing approach, are likely to be swamped by the rotational components. (Note that the author still has reservations, as this theoretical analysis has not been experimentally verified.)

Another consequence of the image projection is that the image flow is invariant under equal scaling of both the position P and the velocity V.

This can be seen by considering tx=xVz/Z-Vx/Z. If P1=(X1,Y1,Z1)=kP= (kX,kY,kZ) and V1=kV then V1z/Z1=Vz/Z and similarly for V1x/Z so tx doesn't change. Similarly for ty

This means that if there is only translational motion of the camera, it isn't possible to tell from the optic flow if an object is close and the camera is moving slowly, or if the object is further away and the camera is moving quickly.

On the other hand, if P is scaled by k then rx and ry are left unchanged because x1=X1/Z1=kX/kZ=X/Z=x. This means that all objects in the viewed scene that are along the vector P will have the same image flow due to rotation, so again it isn't possible to tell if they are close or further away. Combining the translational and rotational motions doesn't help as they are independent, according to a theorem by Euler, which is discusse in [1] Thus it is not possible to extract the altitude from the optic flow.

For a positive translation along the Z axis, the point on the image plane at (Vx/Vz, Vy/Vz) is the Focus of Expansion, (FOE) and lies at the intersection of the image plane with the vector Vz. The optic flow will appear to radiate from this point. [1]

If this point can be identified, (and this may be difficult as the FOE may not actually be within the bounds of the image) it is possible to calculate the time that it will take before a collision with the ground will occur. Because the optic flow is away from (or towards in the case of motion in the negative Z direction) the FOE in all directions, the FOE can be found by finding the minimum (or maximum) in the divergence of the optic flow field.

The divergence of a vector field is given by the vector dot product of the gradient operator with the vector field. If we denote the vector field that represents the optic flow, and which has components u,v in the x and y directions as $\mathbf{A}=(u\mathbf{i}+v\mathbf{j})$, then:

$$\nabla \bullet \mathbf{A} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \qquad \text{Eqn 4}$$

The optic flow itself will be apparently zero at the FOE, but the divergence of the optical flow in a small region around the FOE can be described by Eqn 4.

In the case where the motion is directly along the Z axis the divergence is equal to ?I where I is the identity matrix and ? is a constant that is equal to 1/t where t is the time to collision. [2] It is theoretically possible to calculate this time even given the ambiguity mentioned above, in the distance and velocity, because t is the result of the ratio of the distance and velocity, so the scale factor cancels out.

Another thing to note is that the curl of the optic flow field is zero where it is purely translational. [3] The curl of a vector field is given by the cross product of the gradient operator with the vector field which produces a vector in the Z direction. The magnitude of this vector is a measure of the 'rotational' nature of the field.

$$\left| \nabla \times \mathbf{A} \right| = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \qquad \text{Eqn 5}$$

So theoretically it is possible to use the curl and divergence of A to determine the position of the FOE, and to decide if the optic flow is due to translation or rotation, or both.

The planar version of Green's theorem :

$$\oint_C \mathbf{A} d\mathbf{r} = \iint_R \nabla \bullet \mathbf{A} dx dy = \iint_R \nabla \times \mathbf{A} \bullet \mathbf{Z} dx dy$$

implies that it is possible to calculate the integral of the divergence or curl of a vector field over a surface R, not by applying the differential equations of Eqn 4 and 5, but rather by integrating the vector field around the closed edge C of the surface. The advantage of this is that it avoids the differential. The problem with the differential is that it would be applied to the optic flow field, itself a product of the first order differential of the intensity values of the image, thus resulting in a second order differential of the image. Such second order differentials are notoriously susceptible to noise, and noise is certainly present in the considered video image.

The application of this theorem to this research is still being investigated by the author.

5. Some examples of optic flow patterns and their interpretation

It is important to stress that the following examples have been chosen not because they are typical, but because they show behaviour that the author wants to highlight. Typical flow fields are much harder to interpret, even for a human observer. There is still much work to be done before the author's software can reliably interpret the optic flow. Even these 'chosen best' images are difficult to interpret in isolation, especially those that contain disparate non-zero amounts of translational and rotational component. E.g. mainly a roll, but a small amount of translation as well.

The first optic flow example, Fig. 3 is of motion mainly along the Z axis. The divergence image shows a distinct maximum in the magnitude of the divergence. In the divergence and curl plots, the 'hotter' colors indicate greater magnitude.

The second, Fig. 4 shows what appears to be a translation in the vertical sense, but is actually due to a rotational pitching motion, with perhaps a hint of roll to the left. Apart from the differences in the magnitude of the optic flow, it appears to be quite difficult to discriminate between pitch and rising or falling just from the optic flow, the discussion in section 3 notwithstanding. (This is an issue for future work)

The third, Fig. 5 shows the rotational field and curl due to a roll. The curl is a maximum at the rotational axis of the flow field.

The final flow field, Fig. 6 is due to combined rotation and translation, a roll to right and yaw to left.
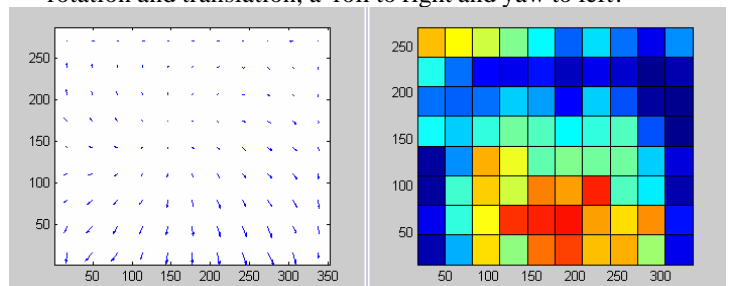

Fig. 3 Optic flow and divergence for forward translation
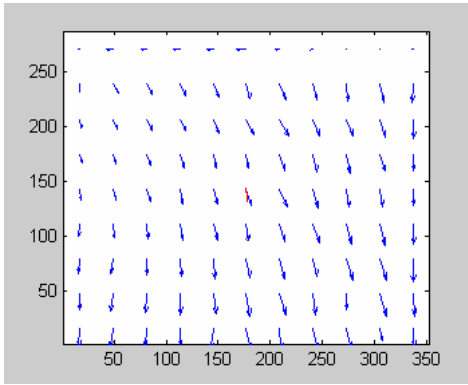
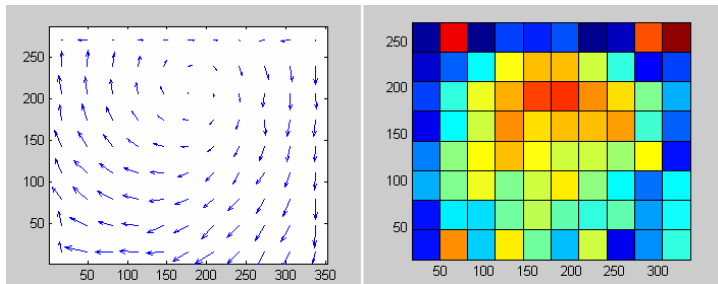Fig. 4 Optic flow from pitching motion

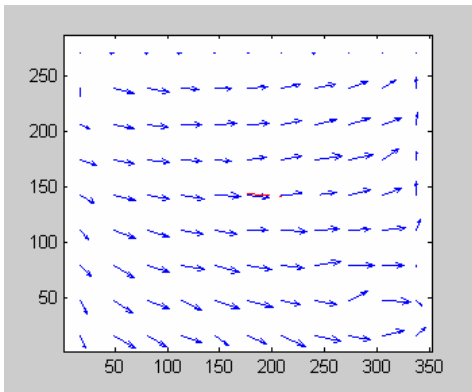Fig. 5 Optic flow and curl for roll to the left

Fig. 6 Combined roll to right and yaw to left

6. Conclusions and Future work

The conclusions that can be drawn thus far are by no means concrete, but it does seem that the optic flow fields indicate quite a lot of useful information about the motion of the UAV. Concerns persist in the author's mind about the accuracy of the calculated optic flow, and about the difficulty of making reliable interpretations from the flow patterns.

Future work includes using the GPS data to quantify the accuracy of the optic flows derived. GPS gives an altitude, ground speed and heading, all of which can be used for verification. Identification and testing of reliable methods to partition the optic flow into translational and rotational components needs to be done. Part of the work will also be towards quantifying the reliability of such interpretations, which is essential because it is a foregone conclusion that some video will not produce reliable optic flow, due to lighting conditions, view angles and texture or lack of it in the view and transmission dropouts even before consideration is given to the accuracy and reliability of the flow fields that the discussed method produces.

7. References

[1]. "A guided tour of computer vision" Vishvjit S.Nalwa Published by Addison-Wesley 1993 ISBN 1-202-54853-4

[2]. "Marco Campani and Allessandro Verri "Motion Analysis from First-order Properties of Optical Flow" Image Understanding Vol 56, no 1, July, 90-107, 1992

[3]. "Vector Analysis and introduction to tensor analysis" Murray R Spiegal. Published by McGraw-Hill

[4]. "Determining Optical Flow" B.K.P. horn, B.G. Schunck, Artificial Intelligence, Vol 17, 185-203, 1981