Accuracy of a Video Odometry System for Trains

Problem presented by

Richard Shenton

Reliable Data Systems

Executive Summary

Reliable Data Systems is developing a video-based odometry system that enables trains to measure velocities and distances travelled without the need for trackside infrastructure. A camera is fixed in the cab, taking images of the track immediately ahead, at rates in the range 25–50 frames per second. The images in successive frames are 'unwarped' to provide a plan view of the track and then matched, to produce an 'optical flow' that measures the distance travelled.

The Study Group was asked to investigate ways of putting bounds on the accuracy of such a system, and to suggest any improvements that might be made. The work performed in the week followed three strands: (a) an understanding of how deviations from the camera's calibrated position lead to errors in the train's calculated position and velocity; (b) development of models for the train suspension, designed to place bounds on these deviations; and (c) the performance of the associated image processing algorithms.

Version 1.2
September 2008

Contributors

Maurice Blount (University of Cambridge)
Paul Dellar (University of Oxford)
Jens Gravesen (Technical University of Denmark)
James MacLaurin (University of Oxford)
Sarah McBurnie (University of Oxford)
Andrew Stewart (University of Oxford)
David Szotten (University of Manchester)
Robert Whittaker (University of Oxford)
Dave Wood (University of Warwick)

Report coordinator

Robert Leese (Smith Institute)

ESGI64 was jointly organised by
Heriot-Watt University
The Knowledge Transfer Network for Industrial Mathematics
The International Centre for Mathematical Sciences

and was supported by
Engineering and Physical Sciences Research Council
The London Mathematical Society
The Institute of Mathematics and its Applications
The European Journal of Applied Mathematics

Contents

1 Introduction

1.1 Background

(1.1.1) In the next generation of railway management systems, such as the European Train Control System (ETCS), there is a trend away from the use of trackside infrastructure to detect the positions and speeds of trains. Reliable Data Systems (RDS) is developing a video-based odometry system that enables a train to measure distances travelled, using a forward-facing camera mounted in the cab. Such a system can report train positions via a radio data link in real time to the signalling control centre, which in turn can provide information to the train on braking points, *etc.*

(1.1.2) In the short term, the benefits of the video system are likely to be in terms of lower costs and higher accuracy of positioning. In the longer term, there are possibilities for allowing closer separation of trains, leading to higher capacity of the rail network.

(1.1.3) Alternative approaches suffer from various shortcomings. Systems that rely on trackside infrastructure are expensive to install and maintain. For systems on the train, there may be limitations on performance in winter conditions, for example due to wheel slip for devices that measure wheel rotations. Inertial methods are expensive. There has been much work done on satellite positioning, but visibility is not consistent (for example in tunnels) and so a secondary sensor system is needed to provide coverage during those periods. The video system being developed by RDS appears to overcome all these problems.

(1.1.4) The overview operation of the video-based system is as follows. The camera mounted in the cab images the track immediately ahead of the train, generally at a frame rate in the range 25–50 frames per second. Each image is 'unwarped', to provide a plan view as if viewed from directly above the track. The unwarped images from successive frames are matched by looking at pixel blocks, to build up an 'optical flow' from one image to the next. This flow provides an estimate of the distance moved between frames.

(1.1.5) Early trials have indicated that this technique is robust, and works effectively at a range of train speeds and in a variety of weather and lighting conditions. However, in order for such a system to be adopted by the industry, it is important to have (very) high confidence that the positioning is accurate to within known error bounds. In ETCS the agreed design requirement is that for a distance travelled s (from some reference point) the accuracy in position shall be better than $\pm\,(0.05s + 5\,\text{metres})$. For estimates of speed, the accuracy requirement is $\pm\,2\,\text{km/h}$ for speeds up to 30 km/h, then increasing linearly up to an accuracy of $\pm\,12\,\text{km/h}$ at a

speed of 500 km/h. Empirically, the accuracy of the current RDS position estimates is better than $\pm 0.025s$, but it is important to provide a rigorous underpinning for such claims.

(1.1.6) The three questions addressed to the Study Group were:

(a) What accuracy claims can be made for the existing system?
(b) What improvements might be made to the existing system?
(c) What accuracy claims can be made for any improvements?

(1.1.7) After a recap of the main sources of error, Section 2 of this report deals with error in the existing system due to vehicle body motion. A linear approximation is developed in detail, along with supporting evidence that indicates the nonlinear corrections are generally small. Section 3 looks at models of the train suspension, in order to construct bounds on the possible variations in camera position and orientation. Section 4 looks at some of the imaging processing issues, and Section 5 summarises the main conclusions.

1.2 Sources of error

(1.2.1) The possible sources of error in the systems are:

(a) **Imaging deficiencies:** these include inaccurate calibration of the camera, errors in the timing of video frames, and motion blur due to finite exposure times. These are not felt to be the major sources of error, and can in any event be addressed through upgrading the camera hardware.
(b) **Vehicle body motion:** trial results indicate that this is the dominant source of error, and is where the Study Group focussed its attention. The position and orientation of the camera will in general be affected by movements in the train suspension. In the current system, the unwarping is fixed, following initial calibration. Therefore small changes in position and orientation will mean that the unwarping is slightly in error, and these errors will propagate to the estimates of position.
(c) **Nonplanar track bed:** the positioning calculations assume that the track bed is planar. In practice, sleepers and ballast are 'bumpy', or there may be uneven snow cover. There may be also be changes in track gradient and occasional sharper raised areas, such as those at level crossings.
(d) **Image processing:** errors can result from the incorrect correspondence of image points (for example caused by lack of detail on the imaged surfaces) and pixel quantisation. The effects of vehicle body motion on the image processing algorithm were also looked at in the Study Group. The need for real-time operation is a constraint on the computational complexity of the image processing.

(e) **Cornering:** errors due to the additional rotational motion when cornering are a particular case of errors due to more general vehicle body motion.

2 Vehicle body motion

2.1 The unwarping transformation

(2.1.1) The camera image and its unwarping are related by a projective transformation. The starting point for assembling this transformation is shown in Figure 1. The coordinate x measures distance along the track, from some fixed point on the ground, with y being in the transverse direction. Let the camera be at position (x_D, y_D), and at a height H above the plane of the track. Three angles describe the orientation of the camera: a declination or 'pitch' θ from the horizontal, a 'yaw' angle ϕ about the vertical, and a 'roll' ψ about the axis of the camera.

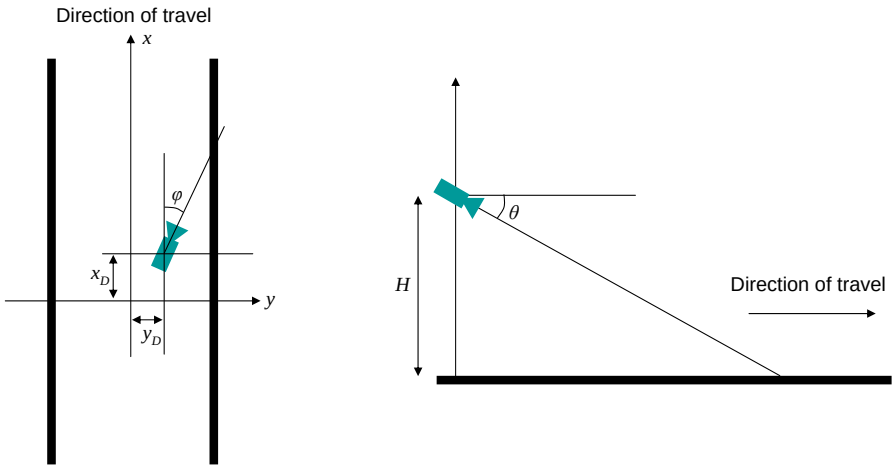

Figure 1: The initial variables used in the analysis of the video system, with the plan view on the left and the side view on the right.

(2.1.2) A general point in the plane of the track has coordinates $(x, y, 0)$. Relative to the position of the camera, its coordinates become $(X, Y, -H)$, where $X = x - x_D$ and $Y = y - y_D$. Transforming into coordinates (ξ, η, ζ) relative to the frame of the camera involves applying three rotation matrices, so that

$$\begin{pmatrix} \xi \\ \eta \\ \zeta \end{pmatrix} = R_1(\psi)\, R_2(\theta)\, R_3(\phi) \begin{pmatrix} X \\ Y \\ -H \end{pmatrix}, \tag{1}$$

where

$$R_3(\phi) = \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{2}$$

$$R_2(\theta) = \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix}, \tag{3}$$

$$R_1(\psi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & \sin\psi \\ 0 & -\sin\psi & \cos\psi \end{pmatrix}. \tag{4}$$

(2.1.3) The coordinates (u, v) in the camera image are related to (ξ, η, ζ) by the projection

$$\begin{pmatrix} u \\ v \end{pmatrix} = f \begin{pmatrix} \eta/\xi \\ \zeta/\xi \end{pmatrix}, \tag{5}$$

where f is the focal length.

(2.1.4) However, the unwarping transformation used by the system assumes that the camera is at a height H_0 above the track, with a pitch of θ_0 and with zero roll and with fixed yaw. We shall take the yaw of the unwarping transformation to be zero, although in practice it is determined by calibrating against the the parallel lines formed by the rails. Coordinates (X^*, Y^*) in the unwarped image are then related to coordinates in the camera image through a rotation $R_0 \equiv R_2(\theta_0)$. Explicitly,

$$\begin{pmatrix} X^* \\ Y^* \end{pmatrix} = -H_0 \left[R_0^{-1} \begin{pmatrix} f \\ u \\ v \end{pmatrix} \right]_{1,2} \Big/ \left[R_0^{-1} \begin{pmatrix} f \\ u \\ v \end{pmatrix} \right]_3, \tag{6}$$

meaning that X^* corresponds to taking the first component in the numerator and Y^* to taking the second.

2.2 Errors introduced by unwarping

(2.2.1) In general we are interested in the effects of small deviations in camera position and orientation from its calibrated position $(\theta, \phi, \psi, x, y, H) = (\theta_0, 0, 0, x_{D0}(t), y_{D0}, H_0)$, where $dx_{D0}(t)/dt = V$. We denote the small deviations by $\delta\theta$, $\delta\phi$, $\delta\psi$, δx_D, δy_D and δH, all functions of time.

Then the perturbed orientation is

$$R = R_1(\delta\psi) R_2(\theta_0 + \delta\theta) R_3(\delta\phi), \tag{7}$$

or equivalently $R = R_0 + \delta R$, where δR may be written in terms of $\delta\theta$, $\delta\phi$ and $\delta\psi$ using (2)–(4). Furthermore, the perturbed camera location is

$$(x_D, y_D, H) = (x_{D0}(t) + \delta x_D\,,\; y_{D0} + \delta y_D\,,\; H_0 + \delta H). \tag{8}$$

(2.2.2) The unwarping transformation then becomes

$$\begin{aligned}\begin{pmatrix} X^* \\ Y^* \end{pmatrix} &= -H_0 \left[(I + R_0^{-1}\delta R) \begin{pmatrix} X \\ Y \\ -(H_0 + \delta H) \end{pmatrix} \right]_{1,2} \Bigg/ \\ &\qquad \left[(I + R_0^{-1}\delta R) \begin{pmatrix} X \\ Y \\ -(H_0 + \delta H) \end{pmatrix} \right]_3 . \end{aligned} \tag{9}$$

(2.2.3) The error in the estimated position that is caused by vehicle body motion comes from comparing (X^*, Y^*) and (X, Y). If we consider two successive time steps, δt apart and label the coordinates of some fixed point on the track at these two times by subscripts 1 and 2, then the actual mean velocity V is equal to $(X_1 - X_2)/\delta t$.[1]

(2.2.4) By linearising (9) around the calibrated camera position, we are able to calculate the leading errors to the velocity estimate:

$$\begin{aligned}\frac{X_1^* - X_2^*}{\delta t} &= V + \delta\dot{x}_D - \frac{\overline{\delta H}}{H_0} V - H_0 \dot{\delta\theta} \\ &\quad + \overline{X} \left(\frac{\dot{\delta H}}{H_0} + \frac{2V}{H_0} \overline{\delta\theta} \right) \\ &\quad + \overline{Y} \left(\sin\theta_0 \, \dot{\delta\psi} - \dot{\delta\phi} - \frac{V}{H_0} \cos\theta_0 \, \overline{\delta\psi} \right) \\ &\quad - \overline{X}^2 \left(\frac{\dot{\delta\theta}}{H_0} \right) \\ &\quad + \overline{X}\,\overline{Y} \left(\frac{\cos\theta_0}{H_0} \, \dot{\delta\psi} \right) . \end{aligned} \tag{10}$$

[1]Note that X is measured relative to the frame of the camera and so as the train moves forwards X decreases, meaning that $X_2 < X_1$; hence V is positive in our expressions.

The corresponding error in the transverse direction is given by

$$\begin{aligned}\frac{Y_1^* - Y_2^*}{\delta t} &= \dot{\delta y_D} - V\overline{\delta\phi} + H\cos\theta_0\dot{\delta\psi} + V\sin\theta_0\overline{\delta\psi} \\ &\quad + \overline{X}\left(\dot{\delta\phi} - \sin\theta_0\dot{\delta\psi}\right) \\ &\quad + \overline{Y}\left(\frac{\dot{\delta H}}{H_0} + \frac{V}{H_0}\,\overline{\delta\theta}\right) \\ &\quad - \overline{X}\,\overline{Y}\left(\frac{\dot{\delta\theta}}{H_0}\right) \\ &\quad + \overline{Y}^2\left(\cos\theta_0\,\dot{\delta\psi}\right). \end{aligned} \tag{11}$$

In these expressions, an overbar indicates the average value across the two timesteps, *e.g.* $\overline{\delta H}$ is equal to $\frac{1}{2}(\delta H_1 + \delta H_2)$, and an overdot indicates a rate of change between the two time steps, *e.g.* $\dot{\delta H} = (\delta H_2 - \delta H_1)/\delta t$. All other terms involving overbars and overdots are defined similarly.

(2.2.5) These errors have been derived from purely geometrical considerations. They are also pointwise estimates, taking no account of any matching of features in successive images, but they are sufficient to make some useful observations.

(2.2.6) Replacing $\overline{X}$ and $\overline{Y}$ with X^* and Y^* in (10) and (11) incurs only $O(\delta^2)$ errors. Thus, given a particular camera offset or motion, we can predict the displacement field between two successive unwarped images. This information can be used in two ways:

(a) We can use the displacement field to estimate the error in the computation of the train velocity by some given algorithm, in terms of the camera motion.

(b) We can attempt to fit the theoretical displacement field to that observed between two images, by varying the coefficients of $\overline{\delta\theta}$ *etc.* This allows most of the artifacts caused by camera motion to be identified and removed. Looking at the terms in (10) that are independent of $\overline{X}$ and $\overline{Y}$, it can be seen that it is not possible to distinguish a change in camera height from a change in speed of the train. Likewise we cannot distinguish $\dot{\delta y_D}$ from $V\overline{\delta\phi}$, but all other camera motions produce independent displacement fields, and so in principle the other coefficients can be identified by fitting.

(2.2.7) To illustrate the possible use of expressions (10) and (11), consider oscillations in camera height around a mean of 3 m; specifically suppose H varies

between 2.95 m and 3.05 m, with a frequency of 2 Hz.[2] The velocity errors when $V = 12$ km/h are shown as a vector field in Figure 2, for points between 5 m and 15 m in front of the train and up to 5 m on either side, and at times when the camera passes through its mean height.

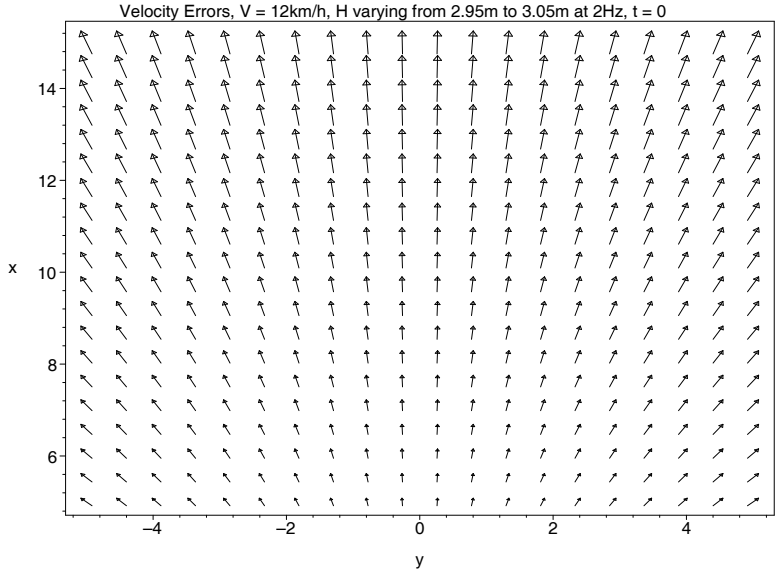

Figure 2: Velocity errors from the linearised analysis when the train velocity is 12 km/h and the camera height oscillates at a frequency of 2 Hz about a height of 3 m with an amplitude of 5 cm.

(2.2.8) Similar plots for velocity errors arising from oscillations in pitch (θ), yaw (ϕ) and roll (ψ) are shown in Figures 3–5. Note that oscillations in θ and H give the same general error pattern, and both have lower errors close to train than further away.

(2.2.9) Train movement around corners and also changes in ground height (for example due to snow) can be expressed as suitable combinations of translations and rotations and are therefore present in the above model.[3]

2.3 Full error treatment

(2.3.1) The linearised analysis of the previous section can be compared with higher-order approximations and with the exact calculation. In this section we investigate the difference between actual error due to vehicle body

[2]We are not suggesting that real oscillations would be so large; in fact, an amplitude of 5 cm at 2 Hz would be extremely uncomfortable for the passenger.

[3]The equation $x_D = x_{D0}(t) + \delta x_D$ is correct to leading order for motion around corners, and therefore can be used in a linearised model.

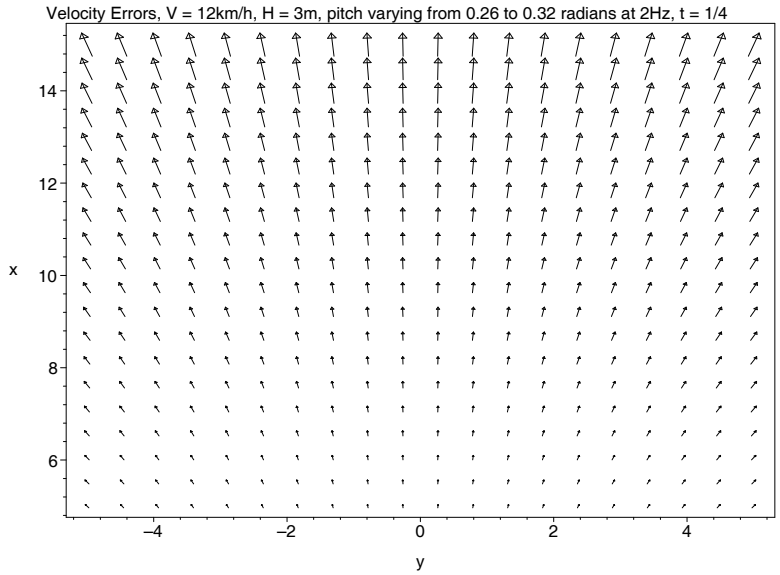

Figure 3: Velocity errors from the linearised analysis when the train velocity is 12 km/h, the camera height is 3 m and the pitch, θ, oscillates between 0.26 and 0.32 radians at a frequency of 2 Hz.

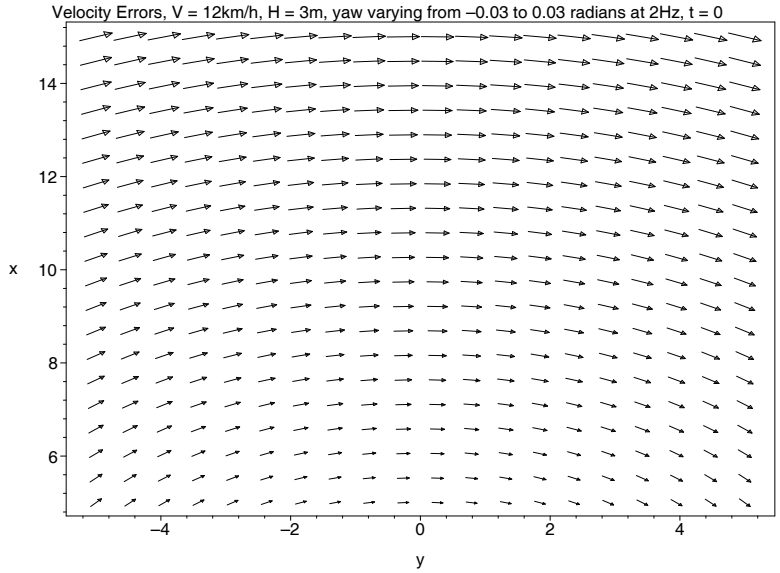

Figure 4: Velocity errors from the linearised analysis when the train velocity is 12 km/h, the camera height is 3 m and the yaw, ϕ, oscillates between -0.03 and 0.03 radians at a frequency of 2 Hz.

motion and the error associated with a linear or quadratic approximation. Suppose that there are two unwarped images between which the train has travelled a distance ΔX. Instead of calculating pointwise er-

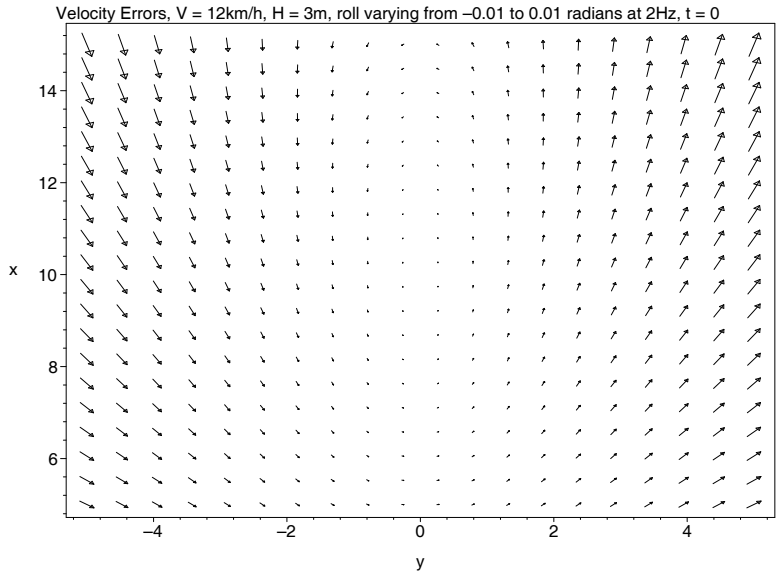

Figure 5: Velocity errors from the linearised analysis when the train velocity is 12 km/h, the camera height is 3 m and the roll, ψ, oscillates between -0.01 and 0.01 radians at a frequency of 2 Hz.

rors, as in the previous section, the two images are matched by finding the translation that minimises the least-squares distance between them.[4] Because of deviations from the calibrated camera parameters, this translation is $(\Delta X + \delta X, \delta Y)$, where $(\delta X, \delta Y)$ is the error caused by vehicle body motion.

(2.3.2) We investigated the differences between the actual error, $(\delta X, \delta Y)$, and linear and quadratic approximations to the error. The parameters in this analysis are deviations $(\delta H_1, \delta\theta_1, \delta\phi_1, \delta\psi_1)$ for the first image, and $(\delta H_1, \delta\theta_1, \delta\phi_1, \delta\psi_1)$ for the second image.

(2.3.3) Visually, we can conveniently investigate the combined effects of pairs of these deviations on a set of axes. It is assumed in the following figures that $H_0 = 3\,\text{m}$, $\theta_0 = 15^\circ$ and $\Delta X = 2\,\text{m}$. Figures 6 and 7 show the inaccuracies that result from linear and quadratic approximations to the error δX, when looking at combined deviations δH_1 and δH_2.

(2.3.4) As another example, figures 8 and 9 show the inaccuracies in linear and quadratic approximations to δX, for deviations $\delta\theta_1$ and $\delta\theta_2$.

[4] In effect, it is being assumed that the image processing algorithms always find the best match between successive frames.

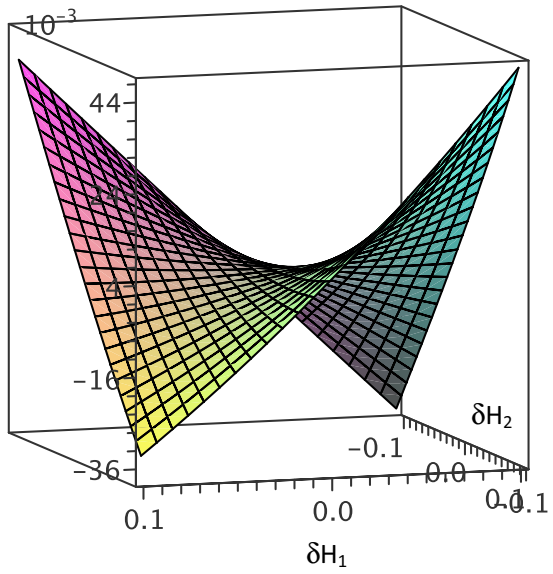

Figure 6: The difference between the actual positioning error δX and a linear approximation, as a function of the deviations in the camera height from the calibrated value in each of the two matched images.

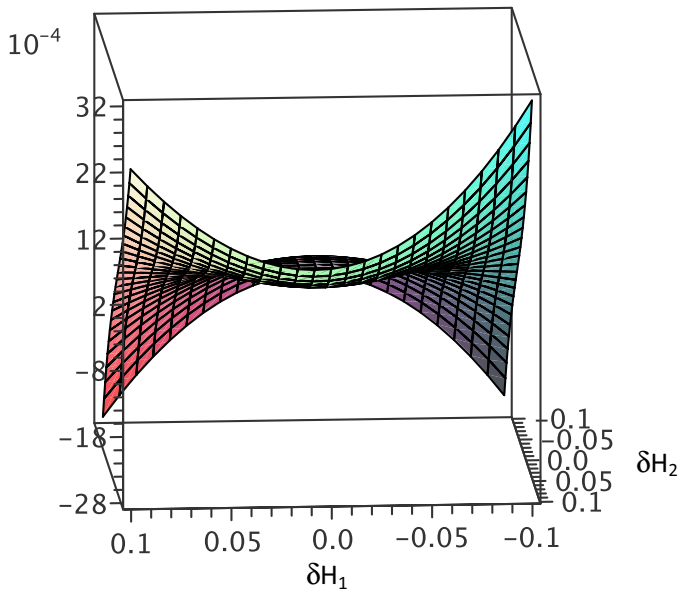

Figure 7: The difference between the actual positioning error δX and a quadratic approximation, as a function of the deviations in the camera height from the calibrated value in each of the two matched images.

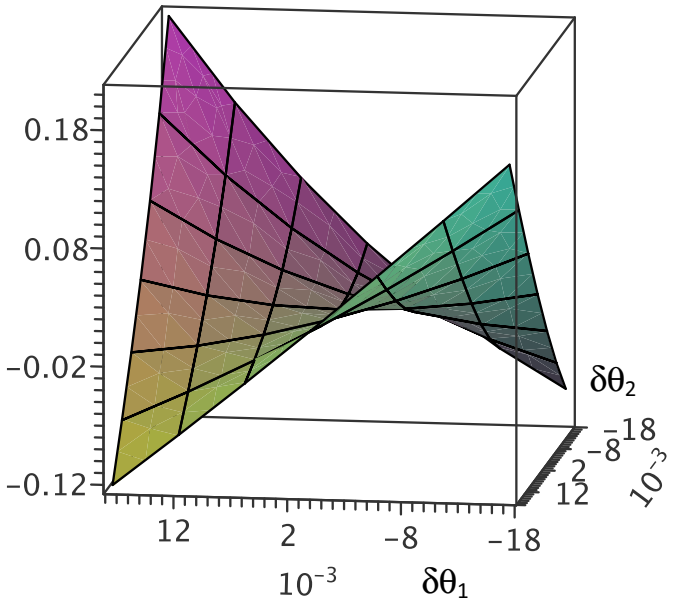

Figure 8: The difference between the actual positioning error δX and a linear approximation, as a function of the deviations in the camera pitch from the calibrated value in each of the two matched images.

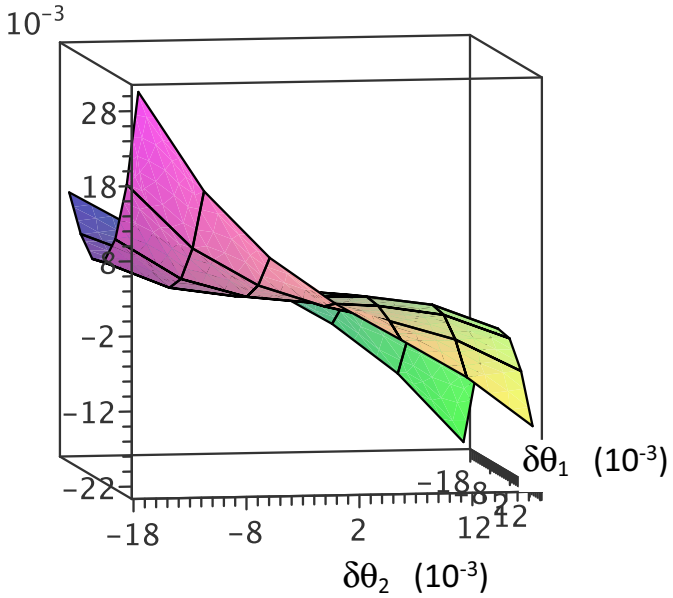

Figure 9: The difference between the actual positioning error δX and a quadratic approximation, as a function of the deviations in the camera pitch from the calibrated value in each of the two matched images.

3 The train suspension

3.1 A static model

(3.1.1) For a static model of the suspension, we considered a fixed chassis, with springs at each corner, on which the carriage is free to move. Assume that the maximum travel of each spring is 10 cm, that the train measures 20 m by 5 m, that the chassis is 1.5 m high, and that the camera is 3 m off the ground. The limits of motion of the springs determine the possible range of camera positions and orientations, and hence the possible positioning errors.

(3.1.2) Calculations of positioning errors were carried out in Maple. The maximum positioning errors arising from vertical motion of the carriage are 50 cm, from pitching are 2 m and from rolling are 3.2 m. They depend strongly on distance ahead of the train. In this model, rolling is the main source of error. The error in velocity is less than 6.4 m/s (or 4 km/h).

3.2 A dynamic model

(3.2.1) A model incorporating linearised dynamics was used to look at oscillations in the pitch of the camera resulting from track vibration. The rail is modelled as an infinite beam with periodic forcing. Two axles (although the model could be readily extended to more) transmit these oscillations to the train via springs and dampers, as shown in Figure 10.

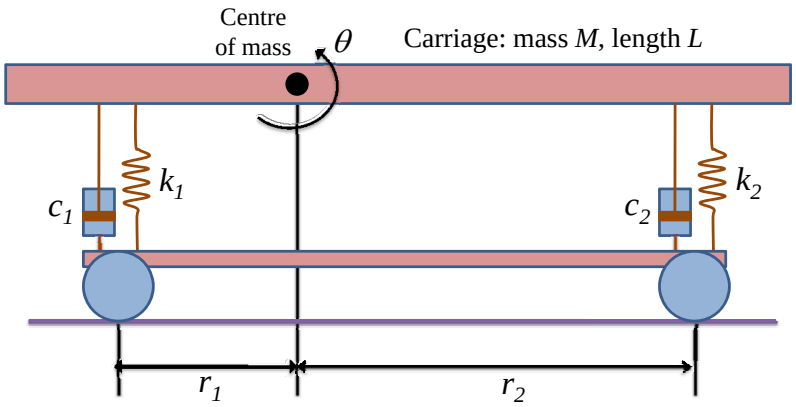

Figure 10: A model of the suspension dynamics, using a spring and damper at each axle.

(3.2.2) Suppose that the spring constants and damping constants are k_1, c_1, k_2 and c_2 (see Figure 10) and that the vertical displacements of the the axles, owing to the periodic forcing of the rails, are u_1 and u_2. Then the vertical

forces, measured positive upwards, that are transmitted to the carriage by the axles are

$$F_1 = k_1(u_1 - u + r_1\theta) + c_1(\dot{u}_1 - u + r_1\dot{\theta}) \quad (12)$$

$$F_2 = k_2(u_2 - u - r_2\theta) + c_2(\dot{u}_2 - u - r_2\dot{\theta})\,, \quad (13)$$

where u is the vertical displacement of the carriage, θ is the pitch of the carriage measured from the horizontal, and r_1 and r_2 are the horizontal distances of the axles from the carriage's centre of mass.

(3.2.3) The equations of motion are $M\ddot{u} = F_1 + F_2$ and $J\ddot{\theta} = -F_1 r_1 + F_2 r_2$, where M and J are the mass and moment of inertia around the pitching axis. Substituting (12) and (13) for F_1 and F_2, and rescaling lengths in units of the carriage length L, we obtain

$$\begin{aligned}\ddot{u} &= \tfrac{1}{2}\omega_0^2(u_1 + u_2 - 2u + (r_1 - r_2)\theta) \\ &\quad + 2\eta\omega_0(\dot{u}_1 + \dot{u}_2 - 2\dot{u} + (r_1 - r_2)\dot{\theta}) \quad (14)\\ \ddot{\theta} &= 6\omega_0^2(r_2u_2 - r_1u_1 + (r_1 - r_2)u - (r_1^2 + r_2^2)\theta) \\ &\quad + 12\eta\omega_0(r_2\dot{u}_2 - r_1\dot{u}_1 + (r_1 - r_2)\dot{u} - (r_1^2 + r_2^2)\dot{\theta})\,, \quad (15)\end{aligned}$$

where

$$\omega_0 = \sqrt{\frac{2k}{M}} \quad \text{and} \quad \eta = c\sqrt{\frac{2}{kM}}\,. \quad (16)$$

(3.2.4) In deriving the right-hand sides of (15), it has been assumed that $J \approx \frac{1}{2}ML^2$, with $L \approx 20\,\mathrm{m}$, and also that k_1, k_2 and c_1, c_2 take common values k and c, respectively. Typical values are $k \approx 3 \times 10^4\,\mathrm{N/m}$ and $c \approx 2 \times 10^4\,\mathrm{Ns/m}$.

(3.2.5) If $r_1 = r_2$ and θ is small then in terms of the rescaled variables the vertical displacement of the camera is approximately $L(u + \frac{1}{2}\theta)$.

(3.2.6) The periodic forcing of the rail is modelled by using u_1 of the form $A\cos(\omega t + \beta)$ and u_2 of the form $A\cos(\omega t)$ in (14) and (15), which then become a pair of coupled linear equations for u and θ. The parameter β allows for a phase shift between the forcing at each axle.

(3.2.7) If the amplitude of the forcing is taken in the range 0.6–1.0 cm (so taking A up to about 0.0005 in terms of the rescaled variables when $L \approx 20\,\mathrm{m}$) then the associated maximum oscillations in camera height are in the range 0.8–1.0 cm and the corresponding oscillation in camera pitch has amplitude of about 0.01 radians.

(3.2.8) There is a connection between the phase difference of forcing at the two axles and the dimensionless **Strouhal number** in fluid dynamics. In the model above, the Strouhal number, St, is $(r_1 + r_2)/(VT)$, where $r_1 + r_2$ is the distance between the two axles, V is the forward velocity of the train, and T is the natural time period of oscillation of the springs. If we assume that the typical distance between successive 'bumps' in the track is much larger than the distance between the axles, then one expects the following qualitatively different regimes in the dynamics: if $\mathrm{St} \ll 1$ then both sets of wheels see bumps in the track at the same time, and the resulting camera displacement is mainly due to changes in u rather than θ; if $\mathrm{St} \approx 0.5$ then there are resonant effect that lead to large variations in θ; and if $\mathrm{St} \gg 1$ then variations in u and θ are both important. The initial numerical solution of (14) and (15) that was carried out at the Study Group indicated that the main contributions came from variations in θ and so the precise relationship between the Strouhal number and train suspension dynamics remains an open question.

4 Image processing effects

4.1 Investigating alternative algorithms

(4.1.1) The current algorithm uses blocks of 16-by-16 pixels to build up a motion field linking one image to the next. Each block in the image is matched to some location in the next image, except where no clear matching can be found. The largest set of similar block displacements is used to estimate the movement of the train from one image to the next.

(4.1.2) Some considerable time was spent at the Study Group looking for systematic error patterns in the motion vectors of the 16-by-16 blocks. None were found, but it should be borne in mind that we were working with JPEG compressed image of the individual camera frames. It is possible that a different result would have been obtained if the raw images had been available.

(4.1.3) Further experiments were carried out using a single larger block, in place of the 16-by-16 blocks, on a sample video clip. In this clip, the train moves of the order of 20 pixels between frames, and so without subpixel resolution the uncertainty in the positioning estimate is about 5%, which is higher than required. Fitting a quadratic to the sum-of-squares difference between images provides good subpixel resolution, as shown in Figure 11.

(4.1.4) It was also found that two independent algorithms, using small (16-by-16) blocks and large blocks gave good agreement. In particular, both algorithms resolve a clear oscillation of approximate frequency 2 Hz in the calculated shift between successive frames (see Figure 12). The fact that

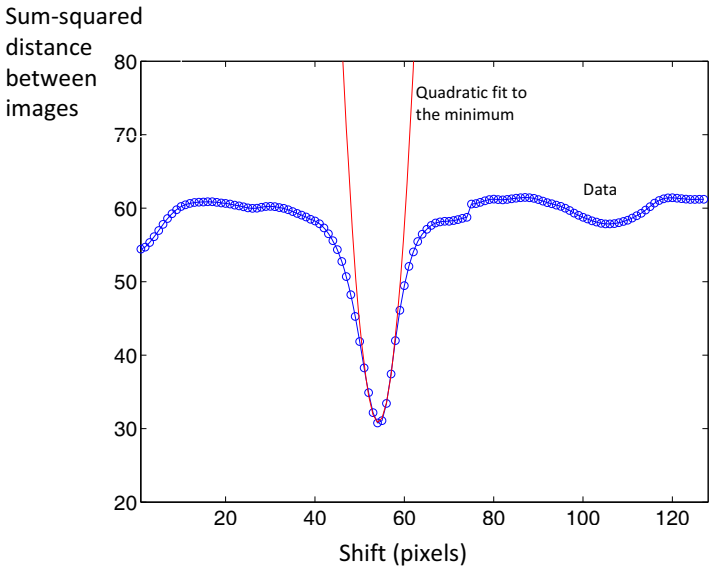

Figure 11: Calculating the shift between images by fitting a quadratic to the sum-of-squares difference.

different algorithms capture the same oscillation suggests that it is not an image processing artefact. Indeed, visually there appears to be appreciable carriage motion in this particular video clip.

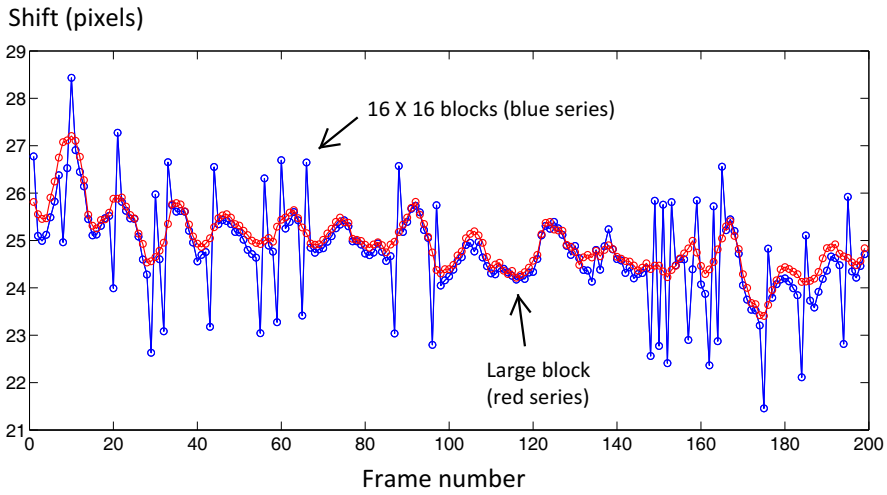

Figure 12: Two independent image processing algorithms, applied to the same video clip. Both resolve a clear oscillation of approximate frequency 2 Hz.

(4.1.5) It should also be noted that the splitting of frames into 16-by-16 blocks is used in MPEG compression. MPEG4 provides further compression by

coding for camera motion, and this may in future make possible the extraction of rigid body motion from the video. See [1] for general reference on MPEG4.

5 Conclusions

5.1 Vehicle body motion

(5.1.1) We carried out a (linearised) calculation of errors due to camera motion. The results may be used either to fit the observed image displacements, so eliminating the error, or to estimate the maximum errors if the camera displacements and rates of change of camera displacements are known. Nonlinear corrections seem to be small.

(5.1.2) Modelling the train suspension provides a means of determining the maximum deviations of camera position and orientation from the calibrated values, and therefore also the magnitude of errors in position and velocity estimates.

5.2 Image processing

(5.2.1) The current algorithm derives an optical flow in terms of 16-by-16 blocks of pixels. It may be possible as an alternative to match the whole image in one go. Experiments with a real video clip indicated that both options successfully capture apparent oscillations in the vehicle body motion.

(5.2.2) Finally, it was noted that the errors are not cumulative, at least when successive train positions are present in a sequence of frames. An error in matching the first and second frames will by removed when matching the second and third frames, so that the cumulative error is just as if the first and third frames had been matched directly. However, 'skipping' frames in this way incurs an image processing loss due to less overlap between the frames being matched. Furthermore, the error field is not uniform across each individual frame (see Figures 2 through 5); errors are generally smaller close to the train.

Bibliography

[1] *The MPEG-4 Book*, Fernando Pereira and Touradj Ebrahimi (Prentice-Hall, 2002)