Premise-Based Estimation

Eyal Shavit

Allan Jepson

Department of Computer Science

University of Toronto

March 15, 1992

Abstract

Instead of computing general descriptions of specific behaviors or specific descriptions of general behaviors, it is often useful to compute specific descriptions of specific behaviors, This is appropriate if for example there are strict time constraints on performance or only a coarse estimate is required. Since little knowledge can be assumed a priori, simplifying assumptions must be made which may or may not be valid. For such an approach to be sound, the computation must be simple, robust to both input and modelling noise, and must facilitate inferences regarding alternative strategies. We propose premise-based estimation as a framework, where premises are made regarding the observed behavior and required description, and the computation is examined within optimal estimation theory. As a demonstration, a coarse description of the motion of arbitrary objects is derived from the premise that they move in a uniform gravitational field. The formulated routine is simple and is shown to have the desirable properties.

1 Introduction

For many complex behaviors, computing a coarse description does not require considering the full detail of change. For example, since the trajectory of the center of mass of an object thrown in the air is parabolic, a routine computing a coarse estimate can ignore detailed object structure, and can rely on the premises of uniform mass distribution and free fall motion which alone provide sufficient constraint. Unfortunately, the center of mass is not available a priori, and the observer must make inferences from the motion of all object points enmasse. Thus, if the object is an "oopee ball" (a ball with non uniform mass distribution) and the observer tracks the ball making the premise that its centoid and center of mass correspond, the observed trajectory appears quite irregular, and definitely not parabolic. The application of such a routine is therefore more involved, and must deal with the fallibility of its underlying premises. This can be due to input noise (deviation from perfect measurements), modelling noise (deviation from the real behavior) or both.

We advocate the framework of *premise-based estimation*, where the computation is based in regularities that are believed to hold for the task and scene, Determining the most useful regularities is based on prior knowledge such as task-specific considerations and the experience of the perceiving system. Determining the regularities most consistent with the sense data can only be determined post hoc. Integrating a priori and a posteriori knowledge is achieved by using the tools of optimal estimation theory [8]. Premise based computation is defined by the lattice theory of perception [7]. Each node of the lattice is defined by a set of regularities termed *premises.* A node or premise set, represents a particular interpretation of the world, which is evaluated by a specialized routine to determine the consistency of its interpretation with the sense data. Knowledge acquired from the concurrent application of many such routines is used to determine the most consistent interpretation.

To demonstrate the principles underlying premise-based estimation, we focus on the task of estimating the motion of objects moving in a uniform gravitational field. These behaviors commonly arise in sports [9], and are of interest since they are performed with accuracy and relative ease. We base estimation on the trajectory of the center of mass of objects. This is appropriate for a large class of behaviors, even where there is deformation, subject to the constraint that the center of mass varies smoothly. It facilitates a description such as trajectory-based classification, determining time-to-collision or computing course-to-intercept [9].

The premise set has two elements; the *motion premise* states that "The motion has a constant acceleration $-g$" which is the known and due to gravity. The *structure premise* is that "The resolution of analysis is λ", constraining the image position of the center of mass at the specified resolution. A review of the possible approaches to this determination, such as feature based techniques, is beyond the scope of our discussion, and the specific method used is left unspecified.

Premise sets previously used in motion estimation,

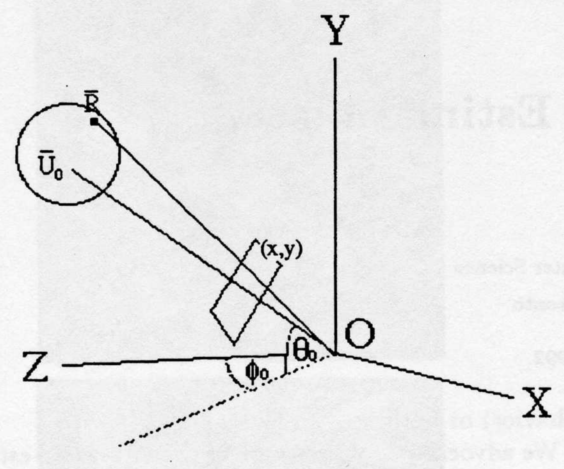

Figure 1: The underlying Geometry (see text).

typically constrain strongly both object motion and structure. The most common premise is rigidity, which constrains both object motion and structure. Structure premises also include articulated models [12], local surface models [11], symmetry [13] or feature point models [3, 5]. Motion premises include smoothness, planarity [6], fixed axis motion [12], pure translation (rotation) [14], constant translation (rotation) [3] or const acceleration [10] (note that this is by no means a complete list of references, for a comprehensive review see [2]). We view the routines designed based on these premises as implementing nodes in the processing lattice. Our premise set strongly constrains object motion but only *weakly* constrains object structure.

2 Design

From the premise set, a two component estimation routine for tracking object motion is formulated. Its batch component (section 2.1) produces an estimate of object position based on an initial viewing direction and two views of the motion. Its recursive component (section 2.2) uses this initial position in estimating subsequent positions based on views generated by an observer tracking the object. The components are applied in succession to the sense data, and from the time evolution of the estimated position all motion parameters can be computed.

The underlying geometry is depicted in figure 1. Let OXYZ denote a fixed inertial space reference with the viewer fixed at its origin. An image is formed along the instantaneous optical axis under perspective projection with the image plane located at unit focal length. Let the direction to a point in space at time index i be the tangents of its polar angles

$$(1) \quad \alpha_i = \tan\phi_i = X_i/Z_i$$
$$\beta_i = \tan\theta_i = Y_i/\sqrt{X_i^2 + Z_i^2}$$

If the optical axis is fixed in direction (α_0, β_0), the image coordinates for an object point $\bar{R}$ are the tangents of its angular disparities relative to the optical axis. Using trigonometry, the image coordinates at time index i are (x_i, y_i) are given by the *projection equation*

$$(2) \quad x_i = \tan(\phi_i - \phi_0) = \frac{\alpha_i - \alpha_0}{\alpha_i\alpha_0 + 1}$$
$$y_i = \tan(\theta_i - \theta_0) = \frac{\beta_i - \beta_0}{\beta_i\beta_0 + 1}$$

The image of an object is analyzed, at the fixed known resolution λ stated by the structure premise, to compute the image position of the premised center of mass. This is termed a *view* of the motion and is denoted by $\bar{m}$. For example, feature based techniques can be used with features defined as 2D blobs, and the computation of $\bar{m}$ accomplished by any one of the methods reviewed in [1].

The trajectory of the center of mass is found by integrating twice with respect to the constant acceleration over the duration of the motion. The acceleration, $\bar{R}'' = [\,0\ ,\ -g\ ,\ 0\,]^T$, is due to gravity, and has a fixed and known direction and magnitude. Denoting the initial or *reference* position and velocity by $\bar{R}_0, \bar{R}'_0$ and letting t_i specify the duration, the motion is contained in a vertical plane and is given by the *trajectory equation*

$$(3) \quad \bar{R}_i = \tfrac{1}{2}\bar{R}''\, t_i{}^2 + \bar{R}'_0\, t_i + \bar{R}_0$$

In the following sections, the motion dynamics (3) and the measurements (2) are used to formulate filter components based in optimal estimation theory [8].

2.1 Batch Component

Define the *state vector* as time index i as $\bar{s}_i = [\,X_i, Y_i, Z_i, Z_0,]^T$ whose elements are the current object position and the depth to the reference point of its trajectory. The *system model* is the motion trajectory (3) and the *measurement model* is the image projection (2). Fixing the optical axis at some known initial direction (α_0, β_0), two views of the motion are accumulated to generate a linear system of equations in the state vector components. The solution $\bar{s}$ along with (α_0, β_0) specify the trajectory completely.

The system of equations is generated as follows; consider three successive positions during the motion, say $\bar{R}_0, \bar{R}_1$ and $\bar{R}_2$. Without loss of generality, select $\bar{R}_0$ as the reference point for the trajectory and fix the optical axis in this direction (α_0, β_0). Denote the views corresponding to the motion, from $\bar{R}_0$ through $\bar{R}_1$ to $\bar{R}_2$, by

$\bar{m}_1, \bar{m}_2$, taken over the intervals t_1 and t_2 measured relative to the reference time t_0. From (3), one positions is a linear combination of the others by t_{ij}

(4) $$\bar{R}_1 = \bar{R}_2 t_1 t_2^{-1} + \bar{R}_0 t_{21} t_2^{-1} + \tfrac{1}{2}\bar{G} t_{21} t_1$$

with $t_{ij} \equiv (t_i - t_j)$.1 From (2), $\bar{m}_i$ and $\bar{R}_i$ are related by

(5) $$X_i + \alpha_i Z_i = 0 \qquad Y_i + \Theta_i Z_i = 0 \qquad \text{where} \quad \alpha_i = \frac{x_i + \alpha_0}{1 - x_i \alpha_0}, \quad \beta_i = \frac{y_i + \beta_0}{1 - y_i \beta_0}, \quad \Theta_i = \beta_i \sqrt{\alpha_i^2 + 1}$$

Algebraically manipulating the equations for $\bar{R}_0, \bar{R}_1, \bar{R}_2$ and the two corresponding views $\bar{m}_1, \bar{m}_2$ yields a *linear* system of equations $\mathbf{A}\bar{s}_2 = \bar{b}$ which depends only on $\alpha_0, \beta_0, \bar{m}_1, \bar{m}_2$

(6)

$$\begin{bmatrix} t_1 & 0 & -\alpha_1 t_1 & -\alpha_{10} t_{21} \\ t_2 & 0 & -\alpha_2 t_2 & 0 \\ 0 & t_1 & -\Theta_1 t_1 & -\Theta_{10} t_{21} \\ 0 & t_2 & -\Theta_2 t_2 & 0 \end{bmatrix} \bar{s}_2 = \begin{bmatrix} 0 \\ 0 \\ -\tfrac{1}{2} g t_{21} t_1 t_2 \\ 0 \end{bmatrix}$$

where $\alpha_{ij} \equiv (\alpha_i - \alpha_j)$ and $\Theta_{ij} \equiv (\Theta_i - \Theta_j)$. A solution exists and is unique when the coefficient matrix $\mathbf{A}$ is non singular. Analytically singularity occurs when the determinant of $\mathbf{A}$ vanishes, that is

(7) $$(\alpha_{20}\Theta_{10} - \alpha_{10}\Theta_{20})\,(t_1 t_2^{-1} - 1) = 0$$

The rightmost term does not vanish as long as measurement times are selected such that $t_1 \neq t_2$. The second term does not vanish if views are such that
Condition 1: *the object is not in the $Z = 0$ plane.*
Condition 2: *the views $\bar{m}_i$ are not collinear.*
The first condition is due to a discontinuity in the projection equation (2), while the second can be seen as follows; substituting from (5) and (1) and rearranging the terms, for $Z_0 \neq 0$, the term vanishes if and only if $\bar{R}_2 \cdot (\bar{R}_1 \times \bar{R}_0) = 0$. That is, if the position vectors during the estimation period are *co-planar*. Since views of the motion are a function of the angular disparities between the position vectors, for the term to vanish the views $\bar{m}_1$, $\bar{m}_2$ must be collinear.

It is easy to see that a solution exists for any motion not entirely in the $Z = 0$ plane and/or does not contain the origin. This can be guaranteed by replacing "bad" views with later ones is the above conditions are not satisfied, updating (6) to reflect the new situation. The existence of "good" views is guaranteed by the planarity of the motion and the non zero curvature of its trajectory. The goodness of solution is expressed by the uncertainty in the state vector. This is defined as the

Models	
system	$\bar{s}'(t) = f(\bar{s}(t), t) + B\bar{u}(t) + G\bar{w}(t)$ $\bar{w}_k \sim \mathcal{N}(\bar{0}; Q)$
measurement	$\bar{m}_k = h(\bar{s}_k) + \bar{n}_k\ ,\ \bar{n}_k \sim \mathcal{N}(\bar{0}; R)$
prior	$E[\bar{s}_0] = \hat{s}_0,\ Cov[\bar{s}_0] = P_0$ $E[\bar{w}\,\bar{n}_k^T] = 0$
Prediction	
state	$\hat{s}_k^- = F_k \hat{s}_{k-1}^+ + \int_{k-1}^k F_\tau B u(\tau)\, d\tau$
covariance	$P_k^- = F_k P_{k-1}^+ F_k^T +$ $\int_{k-1}^k F_\tau G Q G^T F_\tau^T d\tau$
Update	
state	$\hat{s}_k^+ = \hat{s}_k^- + K\left[\bar{m}_k - H(k;\hat{s}_k^-)\hat{s}_k^-\right]$
covariance	$P_k^+ = P_k^-\left[I - KH(k;\hat{s}_k^-)\right]$
gain	$K = P_k^- H^T(k;\hat{s}_k^-)$ $\left[H(k;\hat{s}_k^-)P_k^- H^T(k;\hat{s}_k^-) + R_k\right]^{-1}$ $H(k;\hat{s}_k^-) = \frac{\partial\, h(\bar{s}_k, k)}{\partial \bar{s}}\Big\vert_{\bar{s}_k = \hat{s}_k^-}$

expected value of the state error ([8]), and can be characterized through perturbation analysis. It is known for example [4], that an upper bound on the relative error in $\bar{s}$ can be computed from the condition number $\kappa(A)$ and the relative error in $\bar{b}$. These depend only on the error associated with the known quantities $\alpha_0, \beta_0, \bar{m}_i$ and can hence be given by the routine as part of its computation. This upper bound on the relative error of $\bar{s}$ is then the basis for determining the goodness of the estimate.

2.2 Recursive Component

Extended Kalman filtering was chosen since it allows to handle the nonlinearity introduced by the perspective geometry, while not severely increasing the complexity and cost of the computation. The filter is outlined below, while more detailed filter equations are given in the table below. The state vector is defined as $\bar{s}_i = [X_i, X_0, Y_i, Y_0, Z_i, Z_0]^T$ whose elements are the components of the current and reference object position. The initial estimate $\hat{s}_0$ and its uncertainty P_0, are computed from the state vector of the batch procedure and the initial viewing direction (α_0, β_0) through (1).

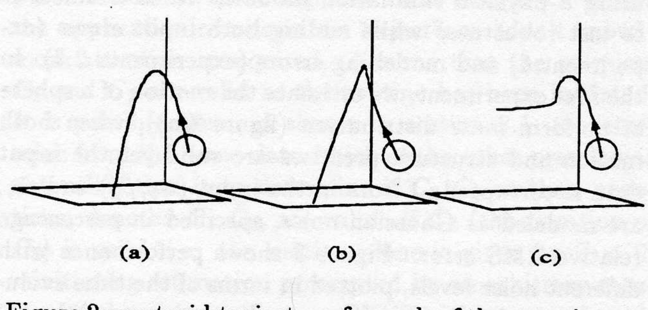

Figure 2: centroid trajectory for each of the experiments described in the text.

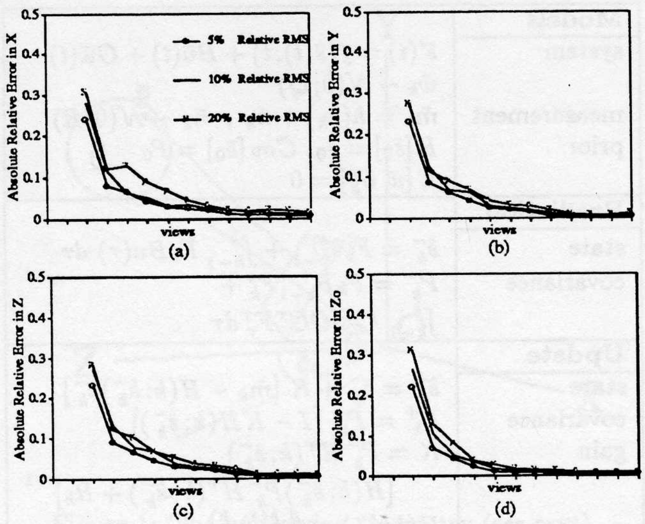

Figure 3: First experiment, both premises are valid (no modelling noise) but noisy input (see text).

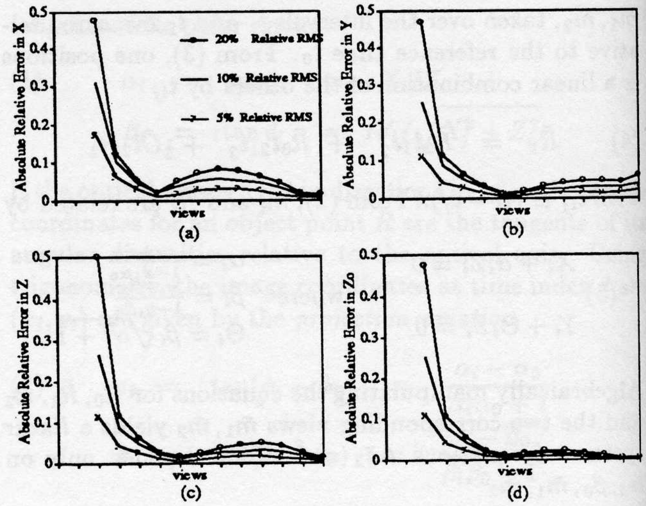

Figure 4: Second experiment, invalid structure premise. Result of 50 Monte-Carlo runs.

The recursive estimation problem is expressed by a linear continuous system model, derived from (3), and a discrete nonlinear measurement model, which is derived from (2). The system model expresses the dynamics of an the center of mass of an object moving under constant gravitational acceleration. The motion is planar and the trajectory parabolic, constrained by the initial velocity and the duration of the motion. The measurement model relates the views, generated by an observer tracking the object, to the corresponding 3D position. Successive views are integrated to estimate the state vector through a set of *prediction* and *update* equations; first $\bar{s}$ is computed forward in time using the system model, and then the views are used to update the prediction through the measurement model. The goodness of the estimate is given by the covariance matrix P which the is computed as an integral part of the estimation.

3 Performance

The routine was applied to object motion generated using a physical simulation module, which enabled us to test robustness while adding both input errors (experiment 1) and modelling errors (experiments 2,3). In the first experiment, we estimate the motion of a sphere of uniform mass distribution (figure 2(a)), when both motion and structure premises are valid yet the input data is corrupted. Errors in the input (α_0, β_0) and $\bar{m}_i$, are modeled as Gaussian noise, specified in percentage relative RMS error. Figure 3 shows performance with different noise levels, plotted in terms of the time evolution of absolute relative error in the estimate of the 3D position X,Y,Z and Z_0. For a tuned filter, implementing pre-run or run-time tuning [8], the errors are given in the square roots of the diagonal terms of the uncertainty matrix P. It is easy to see that there is significant improvement after a small number of views, even in the presence of substantial noise.

In the second experiment we estimate the motion of the same ball, when the motion premise is valid yet the structure premise is not. This modelling error is introduced by varying the mass distribution function of the object within the simulation module. The initial velocity and position are identical to the previous experiment, but the sphere's center of mass was shifted half way along its radius. The trajectory of the centroid is that of an "oopee ball" (figure 2(b)). Modelling error internal to the routine was varied by adding noise to the covariance terms of the filter. Figure 4 shows performance with different levels of error, plotted in terms of absolute relative error in the estimate and time. Performance is significantly improved after the first views of the motion. Subsequent views, about the peak of the trajectory, where the trajectories of the viewed and actual center of mass show maximal discrepancy, cause an increase estimate error. This is not surprising since during this interval, the viewed motion conforms only weakly with that implied by the premise set. The error however remains low as additional views become available, and are substantially lower when *smaller certainty is associated with the input data*. This demonstrates the adaptiveness of the filter and its ability to recover from an erroneous computation. The third experiment involved a sphere when the structure premise is valid yet the motion premise is not. All parameters were set as in

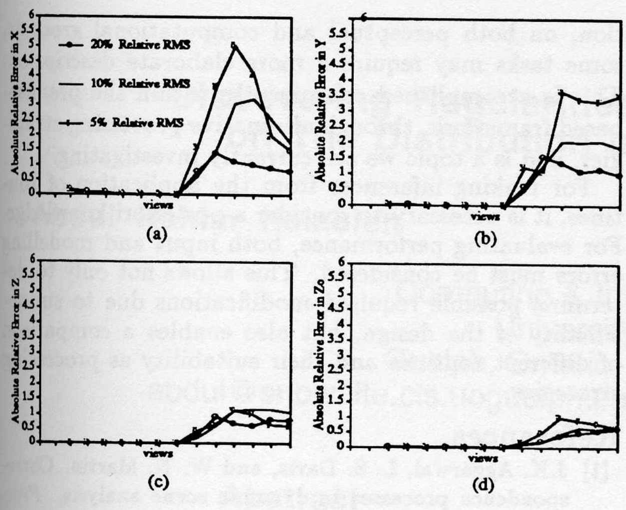

Figure 5: Third experiment, invalid motion premise. Result of 50 Monte-Carlo runs.

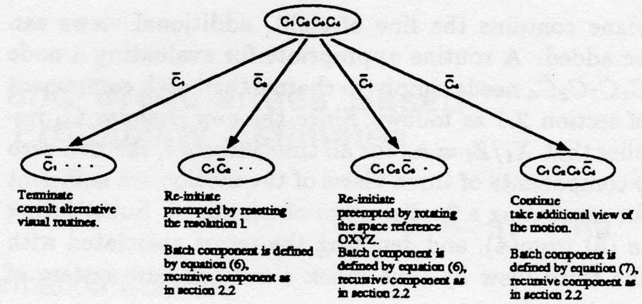

Figure 6: Lattice segment computed by routine.

the first experiment, except that at some point in time, the sphere is subjected to non gravitational acceleration (figure 2(c)). Figure 5 shows performance with different levels of noise, plotted in terms of absolute relative error in the estimate.

In summary, our results demonstrate filter robustness to both image and modelling noise. While currently applied to only a limited class of simulated behaviors, errors are significantly reduced after a small number of views even in the presence of large amounts of noise.

4 Diagnostics

In premise based computation, routines evaluate the consistency of the interpretation suggested by a specific lattice node, which depends on the validity of its premise components as observed in the sense data. The result, as argued in [7], is interpreted as an "accept", "reject" or a "don't know". That is, the interpretation is either accepted or rejected as being consistent with the sense data, or it is stated that there is insufficient evidence for making a decision.

In premise based estimation, this decision step is implemented as a thresholding operation applied to the result of the estimation routine. Thresholding is internal to the filter as real values for the estimated variables are not known. It is based in the state estimate uncertainty and categorized based on two thresholds; where the filter converges, the uncertainty should be below the acceptance threshold, while where it diverges the uncertainty should be above the rejection threshold. When the filter neither converges nor diverges during the estimation period, the associated uncertainty should be in between the thresholds, as no conclusion can be drawn with reasonable certainty. Since behavior is not only a function of design optimality but also depends on filter tunning, the setting of the thresholds for acceptance and rejection should be based on the accumulated experience of the perceiving system in the application of a particular routine. It is often difficult to make a direct inference regarding the validity of any one premise. This is since sufficient constraint is provided by the premise set as a whole, and generally individual premises can not be related in a simple way. One can however produce diagnostics and point to modifications which can the computation.

Generating diagnostics is best understood by through a concrete example. Analytically, the motion estimation routine produces a reliable estimate where both structure and motion premises hold, and further for the initial estimate, where the motion plane is not the $Z = 0$ plane and does not contain the origin. Stated differently, the routine will return a "success" if the following premises are valid

$C_1 \equiv$ "The motion has constant acceleration g".
$C_2 \equiv$ "The resolution of analysis is λ".
$C_3 \equiv$ "The motion plane *isn't* $Z = 0$"
$C_4 \equiv$ "The motion plane *doesn't* contain the origin"

in which case the interpretation denoted $C_1C_2C_3C_4$ is perceived as consistent with the sense data. For tasks with strict performance requirements, parameters such as time-to-collision or course-to-intercept can be computed based on the computed state estimate. Where the application results in an "reject" or "don't know", more attention is required. Assume that the premises C_1 and C_2 are valid as the routine has been initiated. It is possible then, as shown in section 2.1, to determine from the sense data which of the other premises is not valid, that is to determine if C_3, C_4 or their negations, denoted $\bar{C}_3, \bar{C}_4$, are valid. If $\bar{C}_3$ (the negation of C_3) is valid, i.e. the motion plane is given by $Z = 0$, reinitiating estimation preempted by an arbitrary rotation of the viewer reference OXYZ yields a solution. If C_1, C_2 and C_3 are valid yet C_4 is not, i.e. the motion plane contains the line of sight, additional views can be added. A routine appropriate for evaluating a node $C_1C_2C_3\bar{C}_4$ needs simply to change the batch component of section 2.1 as follows; Since the new premise $\bar{C}_4$ implies that $X_i/Z_i = \alpha_0$ for all time indices i, the non-zero y components of three views of the motion are sufficient for generating a 3×3 system of equations. Substituting in (5) from(4), and denoting the terms associated with the third view by time index $i = 3$,the new system of equations is

$$\begin{bmatrix} t_1 & -\Theta_1 t_1 & -\Theta_{10} t_{21} \\ t_1 & -\Theta_2 t_1 & -\Theta_{20} t_{31} \\ 1 & -\Theta_3 & 0 \end{bmatrix} \begin{bmatrix} Y_3 \\ Z_3 \\ Z_0 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} g t_{21} t_1 t_2 \\ -\frac{1}{2} g t_{31} t_1 t_3 \\ 0 \end{bmatrix}$$

When the motion and/or structure premises are not valid, little can be inferred, as C_1 and C_2 are generic to the computational model. Alternative processing strategies, executing simultaneously within the premise-based framework, need be consulted for interpreting the scene. For example, if $\bar{C}_2$ holds, the erroneous resolution of analysis should be varied and estimation reinitiated with a modified structure premise. if $\bar{C}_1$ holds, different regularities of object motion need be found and routines such as those mentioned in section 1 can be consulted.

5 Discussion

Since a successful computation depends on useful and consistent regularities which in turn depend on both task and scene, the use of special purpose routines is a realistic approach to visual processing. This study advocates premise-based estimation as a framework for the design and analysis of such routines, designed from task-determined premises and evaluated using the tools of optimal estimation theory. The issues discussed applies in general to visual processing, and is not limited to motion estimation. Algorithms reported in the literature can be viewed as specialized routines within our framework, which are suitable when the premises they make are valid.

For the design of visual routines it is beneficial to closely examine the task at hand. This can, as was demonstrated for motion estimation, serve to simplify the computation and reduce its dependency on a priori information. The routine formulated in section 2 requires little by way of structure and/or motion information. The necessary structure information is the resolution of analysis, while in terms of motion only few views are needed to produce a good estimate. Its advantage is in its suitability for analyzing a large class of frequently encountered object behaviors, and in the directness and robustness of its computation. Its limitation is in that it yields only a coarse description of behavior. While there support for the relevance of such a description, on both perceptual and computational grounds, some tasks may require a more elaborate description. This is accomplished concurrently within the premise-based framework, through alternative processing strategies, and is a topic we are currently investigating.

For making inferences from the application of routines, it is necessary to consider a posteriori knowledge. For evaluating performance, both input and modelling errors must be considered. This allows not only to determine possible required modifications due to suboptimality of the design, but also enables a comparison of different routines and their suitability as processing strategies.

References

[1] J.K. Aggarwal, L. S. Davis, and W. N. Martin. Correspondence processes in dynamic scene analysis. *Proc. of the IEEE*, Vol. 69, No. 6:562–572, August 1981.

[2] J.K. Aggarwal and N. Nandhakumar. On the computation of motion from sequences of images - a review. *Proc. of the IEEE*, Vol. 76, No. 8, August 1988.

[3] T. J. Broida and R. Chellapa. Estimation of object motion from noisy images. *IEEE Trans. PAMI*, Vol 8, No. 1:90–99, 1986.

[4] G. Dahlquist and A. Bjork. *Numerical Methods*. Printice-Hall, Englewood Cliffs, N.J., 1974.

[5] O.D. Faugeras, F. Lustman, and G. Toscani. Motion and structure from motion from point and line matches. *Proc 1st ICCV*, Vol 9, No. 1:25–34, 1987.

[6] D. D. Hoffman and B. E. Flinchbaugh. The interpretation of biological motion. *Biol. Cybern.*, 42:195–204, 1982.

[7] A. D. Jepson and W. Richards. Perception and perceivers. Department of Computer Science Technical Report RBCV, University of Toronto, Toronto, 1990.

[8] P. S. Maybeck. *Stochastic Models, Estimation, and Control*, volume Vol. 2. Academic Press, 1982.

[9] D. Reagan. The eye in ballgames: Hitting and catching. *Sport en Zien*, pages 7–32, 1990.

[10] B.V.H. Saxberg. Projected free fall trajectories. *Biological Cybernetics*, 56:159–175, 1987.

[11] R.Y. Tsai and T.S. Huang. Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Trans. PAMI*, Vol 6, No. 1:13–26, 1984.

[12] J.A. Webb and J.K. Aggarwal. Structure and motion of rigid and jointed objects. *Artif. Int.*, No. 19:107–130, 1982.

[13] J. Weng, T.S. Huang, and N. Ahuja. 3d motion estimation, understanding and prediction from noisy image sequences. *IEEE PAMI*, PAMI-9(3):370–389, May 1987.

[14] T.D. Williams. Depth from camera motion in a real world scene. *IEEE Trans. PAMI*, Vol 2, No. 6:511–516, 1980.