Recovery of Non-Rigid Motion and Structure

Bradley Horowitz, Alex Pentland
Vision and Modeling Group, The Media Lab
Massachusetts Institute of Technology

Abstract

The elastic properties of real materials provide constraint on the types of non-rigid motion that can occur, and thus allow overconstrained estimates of 3-D non-rigid motion from optical flow data. We show that by modeling and simulating the physics of non-rigid motion we can obtain good estimates of both object shape and velocity. Examples using grey-scale and X-ray imagery are presented, including an example of tracking a complex articulated figure.

1 Introduction

To date almost all research on recovering structure from optical flow has been based on rigid motion, either of surface patches or of 3-D structures. Even schemes which address non-rigid motion, e.g., Ullman's incremental rigidity scheme [1], are usually based on minimizing the deviation from a rigid-body interpretation. Yet non-rigid motion is everywhere: trees sway, flags flap, fish wriggle, arm and leg muscles bunch up, neck and body twist and bend, and cheeks bulge and stretch. One such patch-by-patch recovery of structure is not likely to be either very accurate or robust. Moreover, we want more than a patch-by-patch description, because whole-body motions like bending, twisting and the like are meaningful [4], especially when trying to interpret the actions and gestures of animals and people.

To quote Gibson [5]: "An elastic motion, including that of a walking man with his gestures and facial expressions, could be analyzed into a set of rigid motions of elementary particles if one wished to do so, but it is better thought of in terms of components like bending, flexing, stretching, skewing, expanding, and bulging." Recovering such descriptions is exactly the goal of this paper.

We suggest that the main limitation of previous approaches to non-rigid motion was that non-rigid motion was conceptualized as being completely unstructured. As a consequence all one can say about it is how each point or patch is moving. To describe such completely unstructured motion requires three unknowns per object point, and as a consequence the problem of estimating non-rigid motions becomes badly underconstrained. In fact, however, most real objects are made of approximately elastic materials, and we will show this fact can be used to transform the non-rigid motion problem into an overconstrained problem with a reliable and efficient solution.

The key insight is that the coherent, elastic behavior of real materials implies that non-rigid, whole-body motion can be accurately described with relatively few parameters. The optimal parameterization is obtained from the eigenvectors of the object's corresponding finite element model (FEM) model. These eigenvectors are often referred to as the object's free vibration or deformation modes. The parameterization is unique, and is obtained by a multi-scale orthonormal linear transform (similar to the Fourier transform) that maps the object's point-by-point motion into a coordinate system based on the object's intrinsic deformation modes.

By describing object behavior using a truncated series of vibration/deformation modes one can obtain the best RMS error description possible for a given number of parameters. By varying the number of description parameters (often as a function of the number of sensor measurements available) one can smoothly make the transition from a coarse qualitative description to a finely detailed, accurate description — just as one can smoothly obtain more accuracy by adding more terms to a Fourier series. The important consequence for the problem of recovering non-rigid motion is that the problem can always be made overconstrained by reducing the number of vibration/deformation modes. The limiting case is rigid-body motion, which is equivalent to using only the lowest six vibration/deformation modes.

2 The Finite Element Method

The finite element method (FEM) is the standard engineering technique for simulating the dynamic behavior of an object. Use of the similar but slightly simpler technique of finite differences has become quite popular in machine vision, following the seminal work of Terzopoulos, Witkin, and Kass [6]. In the FEM, interpolation functions are developed that allow continuous material properties, such as mass and stiffness, to be integrated across the region of interest. One major difference between the FEM and the finite difference schemes is that the FEM provides an analytic characterization of the surface between nodes or pixels, whereas finite difference methods do not. All of the results presented in this paper will be applicable to both the finite difference and finite element formulations.

Having formulated the appropriate FEM integrals, they are then combined into a description in terms of discrete nodal points. Energy functionals are then formulated in terms of nodal displacements \( \mathbf{U} \), and the resulting set of simultaneous governing equations is iterated to solve for the nodal displacements as a function of impinging loads \( \mathbf{R} \):
where $U$ is a $3n \times 1$ vector of the $(\Delta x, \Delta y, \Delta z)$ displacements of the $n$ nodal points relative to the object's center of mass, $M$, $C$, and $K$ are $3n$ by $3n$ matrices describing the mass, damping, and material stiffness between each point within the body, and $R$ is a $3n \times 1$ vector describing the $x$, $y$, and $z$ components of the forces acting on the nodes.

When a constant load is applied to a body it will, over time, come to an equilibrium condition described by

$$KU = R \quad (2)$$

This equation is known as the equilibrium governing equation. The solution of the equilibrium equation for the nodal displacements $U$ (and thus of the analytic surface interpolation functions) is the most common objective of finite element analyses. In shape modeling, sensor measurements constitute the recovered shape.

The solution of the equilibrium equation for the nodal displacements $U$ becomes generalized displacements $\hat{U}$:

$$U = \hat{P} \hat{U} \quad (3)$$

Substituting Equation 3 into Equation 1 and premultiplying by $P^T$ transforms the governing equation into the coordinate system defined by the basis $P$:

$$\hat{M}\hat{U} + \hat{C}\dot{\hat{U}} + \hat{K}\ddot{\hat{U}} = \hat{R} \quad (4)$$

where

$$\hat{M} = P^TMP; \quad \hat{C} = P^TCP; \quad \hat{K} = P^TKP; \quad \hat{R} = P^TR \quad (5)$$

With this transformation of basis, a new system of stiffness, mass, and damping matrices can be obtained which has a smaller bandwidth than the original system.

The optimal basis $\Phi$ has columns that are the eigenvectors of $M^{-1}K$ [7]. These eigenvectors are also known as the system's free vibration modes. Using this transformation matrix we have

$$\Phi^T K \Phi = \Omega^2, \quad \Phi^T M \Phi = I \quad (6)$$

where the diagonal elements of $\Omega^2$ are the eigenvalues of $M^{-1}K$ and remaining elements are zero. When the damping matrix $C$ is restricted to be *Euclidean damping*, then it is also diagonalized by this transformation.

The lowest frequency modes are always the rigid-body modes of translation and rotation. The next-lowest frequency modes are smooth, whole-body deformations that leave the center of mass and rotation fixed. Compact bodies—solid objects like cylinders, boxes, or heads, whose dimensions are within the same order of magnitude—normally have low-order modes which are intuitive to humans: bending, pinching, tapering, scaling, twisting, and shearing.

### 4 Recovering Non-Rigid Motion

In this paper we will analyze the case where the object geometry at time $t = 0$ is known, and where object motion is viewed under orthographic projection. The problem, then, is to find the rigid and non-rigid 3-D motions $dU/dt$ that best account for the observed 2-D image velocities. The major difficulty in finding such a solution is that there are $3n$ unknown degrees of freedom in the model and at most $2n$ degrees of freedom in the observations. Thus we must somehow reduce the number of unknowns to obtain a solution.

The modal representation offers a principled, physically-based method for reducing the number of degrees of freedom. Because we know that the elastic properties of real materials imply that the high-frequency modes are (almost) always of low amplitude, we can discard many of these modes without incurring significant error. Further, because the modal representation is frequency-ordered, it has stability properties that are similar to those of a Fourier decomposition. Just as with the Fourier decomposition, an exact subsampling of the data points can never be done by discarding the low-frequency modes. Similarly, irregularities in local sampling and measurement noise tend to primarily affect the high-frequency modes, leaving the low-frequency modes relatively unchanged.

We will therefore pose our problem is the modal coordinate system: the problem is to find the set of 3-D mode velocities $\dot{U}$ that account for the observed 2-D image velocities. If we use the $m$ lowest frequency modes, then there will be only $m$ unknown degrees of freedom in the model and up to $2n$ degrees of freedom in the observations.

Thus by appropriate choice of $m$ the problem can always be made overconstrained.

#### 4.1 Kinematic Solution

We first note that $\phi_i$, the $i^{th}$ column of $\Phi$, describes the deformation the object experiences as a consequence of the modal force $\vec{F}_i$. Or, perhaps more intuitively, $\phi_i$ describes how each of the $n$ nodal points $(x_j, y_j, z_j)$ change as a function of $\dot{u}_i$, the $i^{th}$ mode's amplitude,

$$\phi_i = \left( \begin{array}{c} dx_1 \\ dy_1 \\ dz_1 \\ dx_2 \\ dy_2 \\ dz_2 \\ \vdots \end{array} \right) \left( \begin{array}{c} \frac{d}{dt} u_1 \\ \frac{d}{dt} u_1 \\ \frac{d}{dt} u_1 \\ \frac{d}{dt} u_1 \\ \frac{d}{dt} u_1 \\ \frac{d}{dt} u_1 \end{array} \right) \quad (7)$$

Letting $V$ be the 3-D velocity of each node,

$$V = \left( \begin{array}{c} dx_1 \\ dy_1 \\ dz_1 \\ dx_2 \\ dy_2 \\ dz_2 \\ \vdots \end{array} \right) \left( \begin{array}{c} \frac{d}{dt} \dot{u}_1 \\ \frac{d}{dt} \dot{u}_1 \\ \frac{d}{dt} \dot{u}_1 \\ \frac{d}{dt} \dot{u}_1 \\ \frac{d}{dt} \dot{u}_1 \end{array} \right) \quad (8)$$

we then have that

$$V = \Phi \frac{d}{dt} \hat{U} = \Phi \dot{\hat{U}} \quad (9)$$

Given the 3-D motions of each node, then we can solve for the modal velocities $\dot{\hat{U}}$:

$$\dot{\hat{U}} = \Phi^{-1} V \quad (10)$$
Thus having observed 3-D nodal velocities \( V \), the kinematic solution for the modal amplitudes \( \dot{U}^{t+\Delta t} \) at time \( t + \Delta t \) is simply:

\[
\dot{U}^{t+\Delta t} = \dot{U}^t + \dot{U}' = \Phi^t V^t \Delta t + \dot{U}^t
\]

(11)

The primary limitation of this solution stems from the finite element method's linearization of modes such as rotation. Because these modes are linearized, it is important to limit inter-frame motion to small rotations (less than 10°) and deformations (less than 10% of the object size).

### 4.2 Estimation from 2-D Data

Given the kinematic solution of Equation 11, the remaining problem is to obtain a generalization that uses two-dimensional measurements of optical flow as input data. More concretely, the problem is to estimate the rigid-body motion and non-rigid 3-D object deformation at each subsequent time \( t \) given only noisy estimates of 2-D (orthographically projected) optical flow \( (u_1, v_1) \) at \( m \) image points \((x_1, y_1)\). The image velocities \((x_i, y_i)\) are not assumed to be either dense or uniformly sampled.

This can be accomplished by allocating each of the available optical flow vectors \((u_i, v_i)\) among the nodal points whose image projections are close to \((x_i, y_i)\), the flow vector's image position. This produces estimates of the projected 2-D nodal velocities

\[
V_p = (u_1, v_1, u_2, v_2, \ldots, u_n, v_n)^T
\]

(12)

We define the matrix \( \Phi_p \) similarly, by removing rows of \( \Phi \) that correspond to \( z \)-axis displacements. Note that nodes without nearby optical flow may have no velocity estimate; therefore rows of \( V_p \) and \( \Phi_p \) corresponding to the \( x \) and \( y \) displacements of these nodes are undefined (contain no information) and must also be removed. Similarly, some modes, including translation, scaling and linear shearing along the \( z \) axis, cannot be observed under orthographic projection. Therefore columns of \( \Phi_p \) and rows of \( U \) corresponding to these modes must also be removed.

With these definitions we may now generalize Equation 11 to obtain an estimate of the object's 3-D shape \( \dot{U}^{t+\Delta t} \) at time \( t + \Delta t \) based on the optical flow data. The generalization is simply:

\[
\dot{U}^{t+\Delta t} = \Phi_p^t V_p^t \Delta t + \dot{U}^t
\]

(13)

Equation 13 is underconstrained if all of the modes are present in \( \Phi_p \); however, by discarding a sufficient number of the low-amplitude, high-frequency modes, the estimate can always be overconstrained. Therefore, in practice, \( \Phi_p^t \) is calculated by use of a Moore-Penrose pseudoinverse:

\[
\dot{U} = (\Phi^t \Phi)^{-1} \Phi^t V_p^t \Delta t
\]

(14)

with the columns of \( \Phi_p \) and the rows of \( U \) corresponding to high-frequency modes deleted.

Equation 14 provides us with a least-squares estimate of \( \dot{U} \), the object's rigid motion and non-rigid deformation. It is the best RMS error estimate of the projected rigid and non-rigid motions given the observed optical flow vectors, where the projected mode shapes are described (analytically) by the columns of \( \Phi_p \) and the finite element interpolation functions \( H \).

We have found that 30 deformation modes are adequate to account for most rigid and non-rigid motions, so that only 15 or so independent flow vectors are required per body. In situations with very sparse flow vectors, we can reduce the number of deformation modes still further in order to keep the calculation overconstrained. In the limiting case, we require only three independent flow vectors in order to estimate the six rigid body motions, and four vectors to obtain an overconstrained estimate. Note that this is different than the normal "a views of m points" result in that we are assuming that the initial object geometry is known. Note also the restriction to small inter-frame rotations and deformations.

### 5 A Statistical Evaluation on Synthetic Data

To evaluate the stability and accuracy of the decomposition and estimation process, an experiment was conducted in which randomly selected forces were applied to an elastic spherical body to produce both rigid-body and non-rigid motions. The resulting 2-D optical flow field was then observed, and the rigid and non-rigid motions estimated by use of Equation 14. To make the experiment more realistic, various amounts of uniformly distributed noise was added to the optical flow field before Equation 14 was applied. Each noise condition was repeated with 100 different randomly selected forces and consequent motions. The mean accuracy of the estimation process was then measured.

Figure 1(a) shows the accuracy at estimating rigid-body motion as a function of the signal-to-noise ratio (SNR) of the optical flow field. It can be seen that the accuracy of estimation is linearly related to the SNR of the flow field. The most noisy condition shown here (5 dB SNR) corresponds to approximately 50% added noise.

Again, it can be seen that the accuracy of estimation is linearly related to the SNR of the flow field up to at least 50% noise. The major factor that permits such stable performance is integration of data over the entire body rather than only over a small patch.

### 6 Dynamic Estimation of Rigid and Non-Rigid Motion

In the previous sections we have addressed kinematic estimation, where velocity at only one instant is considered. For time sequences, however, it is necessary to also consider the dynamic properties of the body and of the data measurements. The Kalman filter [8] is the standard technique for obtaining estimates of the state vectors of dynamic models, and for predicting the state vectors at some
later time. Outputs from the Kalman filter are the optimal (weighted) least-squares estimate for non-Gaussian noises.

The first use of Kalman filtering for motion estimation was by Brodia and Chellappa [9], who presented a careful evaluation of the approach. Work by Faugeras, Ayache and their colleagues, and more recently many others, has thoroughly developed the subject [10; 11]. In this section we will develop a Kalman filter that estimates position and velocity for the finite element modal parameters. We will then show that this particular type of Kalman filter is mathematically equivalent to time integration of the FEM governing equation for appropriate choices of mass \( M \) and stiffness \( K \). That is, the Kalman filter may be viewed as a simulation of the model’s behavior, with the observed optical flow acting as guiding “forces.”

6.1 The Kalman Filter

Let us define a dynamic process

\[
\dot{X} = AX + Ba
\]

and observations

\[
Y = CX + n
\]

where \( a \) and \( n \) are white noise processes having known spectral density matrices. Then the optimal estimate \( \hat{X} \) of \( X \) is given by the following Kalman filter

\[
\dot{\hat{X}} = A\hat{X} + K_f(Y - CX)
\]

with correctly chosen Kalman gain matrix \( K_f \).

6.1.1 The Kalman Gain Factor

The gain matrix \( K_f \) in Equation 17 minimizes the covariance matrix \( P \) of the error \( e = X - \hat{X} \). Assuming that the cross-variance between the system excitation noise \( a \) and the observation noise \( n \) is zero, then \( K_f = PC^T \Lambda^{-1} \), where the \( n \times n \) observation noise spectral density matrix \( \Lambda \) must be nonsingular [9]. Assuming that the noise characteristics are constant, then the optimizing covariance matrix \( P \) is obtained by solving the Riccati equation

\[
0 = \dot{P} = AP + PA^T - PC^T \Lambda^{-1} CP + B A B^T
\]

where \( \Lambda \) is the \( n \times n \) spectral density matrix of the acceleration noise \( a \).

6.1.2 Estimation of Displacement and Velocity

In the current application we are primarily interested in estimation of the modal amplitudes \( \dot{U} \) and their velocities \( \dot{V} = \dot{U} \). In state-space notation our system of equations is

\[
\begin{bmatrix}
\dot{\hat{U}} \\
\dot{\hat{V}}
\end{bmatrix} =
\begin{bmatrix}
0 & I \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
\hat{U} \\
\hat{V}
\end{bmatrix} + \begin{bmatrix}
0 \\
1
\end{bmatrix} a
\]

where \( a \) is a noise vector due to nodal accelerations. The observed variable will be the 2-D nodal velocities \( V_p \), and from Equation 13 we have that

\[
\dot{V}_p = (\Phi_F/\Delta t) \dot{U} + n
\]

where \( n \) is a vector of the observation noise. The Kalman filter is therefore

\[
\begin{bmatrix}
\dot{\hat{U}} \\
\dot{\hat{V}}
\end{bmatrix} =
\begin{bmatrix}
0 & 1 \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
\hat{U} \\
\hat{V}
\end{bmatrix} +
\begin{bmatrix}
K_{f1} \\
K_{f2}
\end{bmatrix}
\begin{bmatrix}
V_p - [(\Phi_F/\Delta t) \dot{U}] \\
0
\end{bmatrix}
\]

\[
(21)
\]

where \( K_{f1} \) and \( K_{f2} \) are the Kalman gain matrices for velocity and acceleration, respectively.

We will assume that \( n \) and \( a \) originate from independent noise with standard deviations \( \sigma_n \) and \( \sigma_a \) respectively. As each point-wise measurement error is spread to the various modes by \( \Phi_F/\Delta t \), it is reasonable to choose \( \Lambda' = \sigma_a \Phi_F^2/\Delta t \). Given \( \Lambda' \) and \( \Lambda \), we may then determine the Kalman gain matrices, which for this simple noise model are

\[
K_{f1} = \left( \frac{\sigma_a}{\sigma_n} \right)^{1/2} \Phi_F^{-1} \Delta t
\]

and

\[
K_{f2} = \left( \frac{\sigma_a}{\sigma_n} \right)^{1/2} \Phi_F^{-1} \Delta t
\]

Substituting this result into Equation 21 we obtain

\[
\begin{bmatrix}
\dot{\hat{U}}^{+1} \\
\dot{\hat{V}}^{+1}
\end{bmatrix} =
\begin{bmatrix}
\dot{U} + (\frac{\sigma_a}{\sigma_n})^{1/2} \Phi_F^{-1} \Delta t (V_p - (\Phi_F/\Delta t) \dot{U}) \\
(\frac{\sigma_a}{\sigma_n})^{1/2} \Phi_F^{-1} \Delta t (V_p - (\Phi_F/\Delta t) \dot{U})
\end{bmatrix}
\]

\[
(22)
\]

We can now formulate the displacement prediction at time \( t + \Delta t \). Each mode is independent within this system of equations, and so we may write the Kalman filter for each of the separate modes:

\[
\dot{\hat{U}}^{+1} = \hat{U} + d_1 \dot{\hat{U}} + d_2 (\dot{\hat{U}} - \hat{U})
\]

\[
(23)
\]

which is exactly the central-difference update rule for direct time integration of the finite element governing equations, with “loads” \( \dot{U}_i = \dot{v}_p \Delta t, d_1 = \Delta t, d_2 = 2\Delta t^2/m_i = (a/n) \Delta t^2 + (2a/n) \Delta t, \) and \( \dot{v}_p \) are the elements of \( V_p = \Phi_F^2 V_p \). The equivalence between these Kalman filter equations and time-integration of a finite-element governing equation provides an intuitive interpretation of the Kalman filter. In essence, it is integrating optical flow data over space by modeling it using the low-frequency, whole-body mode shapes, and integrating over time by use of a mass matrix that gives “inertia” to the estimates.

7 An Example Using Synthetic Data

Rotational dynamics and non-rigid dynamics are nonlinear problems that are linearized by the FEM. As a consequence, the Kalman filter developed here may be more properly considered an extended Kalman filter and, despite the well-known stability and accuracy of the FEM, there is no proof of convergence and bias-free behavior. We therefore have evaluated the stability and accuracy of our formulation using synthetic data.

Figure 2 shows an example of tracking both rigid and non-rigid motion. The top row of Figure 2 shows four frames from a 30 frame image sequence of a rotating, translating, and deforming solid. In this example the initial position, shape, and (linearized) velocity of the object was assumed known. Exact optical flow data from this sequence...
inter-frame time step was 0.1 seconds, and the parameters are assumed to be relatively rare). Noise. Optical flow SNR is 10 dB.

ber, for two ratios of acceleration noise to optical flow field. In this example the rigid-body motions were tracked with an error of 29.1 dB SNR (approximately 3.5% error). The error in both rigid-body and non-rigid modes was 18.5 dB SNR (approximately 11.9% error).

The major source of error in the non-rigid modes was introduced by a single ill-conditioned mode, i.e., a mode for which even large deformations cause small only 2-D motions. In Figure 2, for instance, even though the amount of bending perpendicular to the image plane is almost 5% in error the tracking object appears nearly identical to the original object. Such ill-conditioned modes can be detected by examination of the columns of \( \mathbf{A} \), although this was not done in this experiment.

7.1 A Statistical Evaluation

The previous examples have shown that our formulation can produce accurate estimates of motion, but they do not allow evaluation of either convergence or stability. To evaluate these properties, we followed the methodology of Brodie and Chellappa [9], and constructed an experiment in which there were large errors in the initial velocity estimates. This condition is equivalent to the case in which a very large acceleration “spike” produces a large inter-frame change in system velocities. Following this acceleration spike, the behavior of the Kalman filter over successive frames was observed to determine whether or not the Kalman estimates would converge rapidly to the correct value.

In our experiment noisy motion estimates for 100 image sequences were used as input to Equation 23. The motion estimate noise level averaged 20 dB (i.e., the noise magnitude was 10% of the flow vector magnitude). The mean velocity for each mode (including rigid-body modes) was approximately 5 cm/second. The inter-frame time interval was 0.25 seconds. In each trial the initial estimate of each mode's velocity was zero so that the mean initial error was 5 cm/second for each mode. This condition is equivalent to applying an acceleration of 20 cm/second*2 to a resting system between the 0th and 1st frames of the image sequence.

The experiment was repeated with two separate measurement/acceleration noise models, one with \( a/n = 2 \) (large accelerations are common) and one with \( a/n = 0.5 \) (large accelerations are uncommon). The upper curve in Figure 4 shows the estimates convergence with the \( a/n = 0.5 \) model, the lower curve shows convergence with \( a/n = 2.0 \). The error bars show the standard deviation of the 100 separate estimates at each frame number. Note that although the mean error goes very nearly to zero, in individual trials the errors were in the range of ±1%.

As can be seen from Figure 4, stable and accurate convergence was achieved in both cases. All modes, including rigid-body modes, behaved in a very similar manner.

8 Examples Using Real Data

Figure 3 shows an example of recovering non-rigid motion from optical flow data and Equation 23. The 3-D shape and motion of the heart ventricle was tracked over time using this optical flow data. The computation started with an initial 3-D model of the ventricle, shown overlayed on frame 1 of Figure 3. Equation 23 was used to estimate the 3-D rigid and non-rigid motion of the ventricle at each time step. The resulting rigid and non-rigid motions are shown by the wireframe model overlayed on the original X-ray imagery. Execution time was approximately one second per frame on a standard Sun 4/330.

8.1 Constrained motion

In many cases the observed motion is known to be constrained, for example, by gravity or by a hinge or other attachment. Such constrained motion adds a bias or control term to Equation 15, but as long as it varies sufficiently slowly with respect to the Kalman filter's sampling rate it does not otherwise affect the convergence or stability of the estimator [8]. We may therefore hope to use the Kalman filter of Equation 23 to track the rigid and non-rigid behavior of constrained objects as well free-moving objects.

In the Kalman state equations such forces appear as a constant or slowly-varying acceleration bias, i.e., Equation 19 becomes

\[
\dot{\mathbf{U}} = \dot{\mathbf{V}}, \quad \dot{\mathbf{V}} = \mathbf{R}^x + a
\]

where \( \mathbf{R}^x \) is a vector describing the load exerted on each nodal point by all active constraints (see reference [14] for details of the constraint system).

Given a \emph{priori} knowledge of such a motion constraint, we can compensate for the contribution of that constraint to the state equations and then estimate motion as previously. The simplest way to accomplish this is to modify Equation 23 to account for this new term:

\[
\hat{\mathbf{v}}_{i+1} = \hat{\mathbf{v}}_i + d_i \mathbf{1} + d_2 (\hat{\mathbf{r}}_i - \hat{\mathbf{r}}_i) + 2\Delta t^2 \hat{\mathbf{r}}_i / \hat{\mathbf{m}}_i
\]

where \( \hat{\mathbf{m}}_i = \rho \sum_j |\phi_{ij}| \) is an estimate of the \( i \)th mode's generalized mass, parameterized by \( \rho \), an estimate of the object's density.

Figure 5 illustrates a relatively complex example of tracking an object in which motion is a priori known to be constrained to certain part junctions (joints). This figure shows three frames from a twelve image sequence of a well-known tin woodsman caught in the act of jumping. Despite the limited range of motion, this example is a difficult one because of the poor quality optical flow, due to pronounced highlights on thighs and other parts of the body.

![Image](image-url)
9 Summary

We have introduced a precise, physically-correct model of elastic non-rigid motion. This model is based on the finite element method, but decouples the degrees of freedom by breaking down object motion into rigid and non-rigid vibration or deformation modes. Because of the intrinsic elastic properties of real materials, it can be shown that the high-frequency modes in this representation rarely have significant amplitude, so that they may be discarded without introducing undue error.

The result is an accurate representation for both rigid and non-rigid motion that has greatly reduced dimensionality, capturing the intuition that non-rigid motion is normally coherent and not chaotic. Because of the small number of parameters involved, we have been able to use this representation to obtain accurate, overconstrained estimates of both rigid and non-rigid motion.

We have also shown that these estimates can be integrated over time by use of an extended Kalman filter, resulting in stable and accurate estimate of both 3-D shape and 3-D velocity. The formulation was then extended to include constrained non-rigid motion. Examples of tracking single non-rigid objects and multiple constrained objects were presented.

References


