Pose Estimation by Fusing Noisy Data of Different Dimensions

Yacov Hel-Or and Michael Werman

Abstract— **A method for fusing and integrating different $2D$ and $3D$ measurements for pose estimation is proposed. The $2D$ measured data is viewed as $3D$ data with infinite uncertainty in particular directions. The method is implemented using Kalman filtering, it is robust and easily parallizable.**

Keywords— **sensor fusion, Kalman filter, pose estimation, model based.**

In model-based pose determination the position of a known object is determined from different types of surface measurements (for reviews see [1], [2], [3]). Usually feature points such as maximum curvature, segment endpoints and corners are measured. The aim of this paper is to determine the correct rigid transformation (translation and orientation) of the model points to the measured points where the measured data is not exact. This problem is known as *absolute orientation* in photogrametry (for a review see [4]) and is classified into two major categories according to the type of measurements:

1. $3D$ to $3D$ correspondence: both model and measurements supply information about the $3D$ location of features (measurements from range data, stereo, etc.).
2. $2D$ to $3D$ correspondence: the model is $3D$ while the available measurements supply projected $2D$ information. The projection can be perspective or orthographic.

Methods to compute the absolute orientation have been presented, most of which use least-square techniques in either closed (e.g. [5], [6], [7], [8], [9]) or iterative form (e.g. [10], [2], [4], [6], [11]). However, each method can be applied to only one of the categories described above.

In this paper we suggest a uniform framework to compute the absolute orientation, where the measured data can be a mixture of $2D$ and $3D$ information. Unifying the different types of measurements is done by associating an uncertainty matrix with each measured feature. Uncertainty depends both on the measurement noise and on the type of measurement. A $2D$ measurement is a projection (perspective or orthographic) onto a $2D$ plane and we regard it as a measurement in $3D$ with infinite uncertainty in the direction of the projection. Therefore, the dimensionality of the measurements is encoded in the covariance matrix. This representation unifies the two categories of the absolute orientation problem into a single problem that varies only in the uncertainty values associated with the measurements. With this paradigm we obtain a uniform mathematical formulation of the problem and can fuse different kinds of measurements to obtain a better solution. The algorithm we describe has additional advantages of supplying a certainty measure of the estimate, enabling an efficient matching strategy and allows simple parallelization.

A *model* M of a $3D$ object is represented by a set of points:

$$M = \{\mathbf{u}_i\} \quad ,$$

where $\mathbf{u}_i$ is a 3 dimensional object-centered vector associated with the i^{th} point.

A *measurement* of a $3D$ object is represented by M' which, similar to the model representation, is a collection:

$$M' = \{(\hat{\mathbf{u}}'_j, \Lambda_j)\} \quad .$$

$\hat{\mathbf{u}}'_j$ - is a noise-contaminated measure of the real location-vector $\mathbf{u}'_j$ associated with the j^{th} measured point and is represented in a viewer-centered frame of reference.
Λ_j - is the covariance matrix depicting the uncertainty in the sensed vector $\hat{\mathbf{u}}'_j$. We do not constrain the dimensionality of the measured data but allow it to be $3D$ (stereo, range finder etc.), or $2D$ (orthographic or perspective projection).

A *matching* between the model M and the measurement M' is a collection of pairs of the form

$$matching = \{\mathbf{u}_k, (\hat{\mathbf{u}}'_k, \Lambda_k)\} \quad ,$$

which represents the correspondence between the model points to the measured points. For simplicity we denote a model point and its matched measurement with the same indices.
The problem:
Given a model M, a measurement M' and a matching as above, estimate a transformation T which optimally maps the points $\mathbf{u}_i$ of the model onto the corresponding measured points $(\hat{\mathbf{u}}'_i, \Lambda_i)$. The estimated transformation T describes the position of the measured object M' in the $3D$ scene.

The method described below fuses the information from all the measured points and estimates the transformation T by incremental refinement using Kalman-filter tools. At each step a matched pair is introduced and an updated solution is produced.

As previously noted, in our approach, an uncertainty matrix must be evaluated for each and every synthesized point feature. That is, each extracted feature is associated with both, a set of estimated parameter values and an uncertainty matrix associated with these values. This uncertainty is derived from several factors:

- Uncertainty due to measurement noise (e.g. digitization, blurring and chromatic aberrations).
- Uncertainty dependent upon the feature detection process. For example, a detected end-point of a line segment will have a low positional uncertainty in the direction perpendicular to the line segment and a high uncertainty in its direction.
- Uncertainty due to the lack of information caused by projections.

In this paper, we will not deal with the modeling of the measurement noise but we present a unified representation of the measured data. We separate our class of measurements into two categories:

$3D$ measured data:
The simplest case is that of a point $q \in M'$ presented by the pair:

$$q = [\hat{\mathbf{u}}', \Lambda] \quad ,$$

where $\hat{\mathbf{u}}' = (\hat{x}', \hat{y}', \hat{z}')$ is the measured location vector and Λ is its uncertainty.

$2D$ projected data:
When the measurements are obtained using a projection, we

Y. Hel-Or and M. Werman are with the Institute of Computer Science, The Hebrew University of Jerusalem, 91904 Jerusalem, ISRAEL; email toky@cs.huji.ac.il

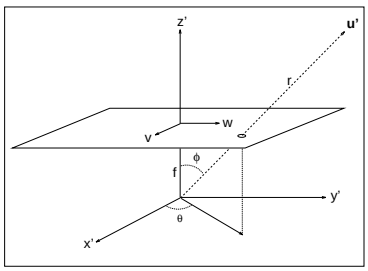

Fig. 1. A perspective projection of the point $\mathbf{u}'$ into an image plane (v, w). The point can be represented in either, a Cartesian system (x', y', z') or a spherical system (r, ϕ, θ).

describe the measured data as a measurement in $3D$ where the uncertainty in the direction of projection is infinite. Assume the measurements are performed on the image plane using the coordinate system (v, w):

$$proj(q) = [(\hat{v}, \hat{w}), \Sigma_{vw}] \quad ,$$

where Σ_{vw} is a 2×2 covariance matrix describing the uncertainty of the measurement $(\hat{v}, \hat{w})$.

In the case where the projection is along the z-axis (orthographic) we represent this data as:

$$q = [(\hat{v}, \hat{w}, \hat{z}'), \begin{pmatrix} \Sigma_{vw} & 0 \\ 0 & \infty \end{pmatrix}] \quad ,$$

where $\hat{z}'$ is any estimate of the z' coordinate.

In the case of a perspective projection, modeling of the uncertainty is a little more complicated. Assume the origin of the viewer-centered frame of reference is at the focal point as shown in Figure 1 and the focal length is f.

We aim to transform the measurement given in the image-plane coordinate system into a representation in the Cartesian system $\mathbf{u}' = (x', y', z')$.

Considering the spherical coordinate system (r, ϕ, θ) (Figure 1). The vector $(\hat{v}, \hat{w})$ determines the angular coordinates (ϕ, θ) but leaves the value of r undetermined:

$$\hat{\phi} = \arctan(\frac{\sqrt{\hat{v}^2 + \hat{w}^2}}{f}) \quad ; \quad \hat{\theta} = \arccos(\frac{\hat{v}}{\sqrt{\hat{v}^2 + \hat{w}^2}}) \quad .$$

Additionally, the uncertainty of $(\hat{v}, \hat{w})$ is translated into a covariance matrix in the (ϕ, θ) system as follows:

$$\Lambda_{\phi\theta} = \left(\frac{\partial(\phi, \theta)}{\partial(v, w)}\right) \Sigma_{vw} \left(\frac{\partial(\phi, \theta)}{\partial(v, w)}\right)^t \quad ,$$

where $\frac{\partial(\phi,\theta)}{\partial(v,w)}$ is the Jacobian of the transform from (v, w) to (ϕ, θ), and the derivative is taken at point $(\hat{v}, \hat{w})$. The transformation into spherical coordinates, as an intermediary stage, allows a simple representation of the measurement in $3D$: $q = [(\hat{r}, \hat{\phi}, \hat{\theta}), \Lambda_{r\phi\theta}]$, where

$$\Lambda_{r\phi\theta} = \begin{pmatrix} \infty & 0 & 0 \\ 0 & & \\ 0 & & \Lambda_{\phi\theta} \end{pmatrix}$$

and $\hat{\phi}$, $\hat{\theta}$, $\Lambda_{\phi\theta}$ are the expressions described above. $\hat{r}$ is unknown but an estimation of $\hat{r}$ will be chosen as is explained later in this section.

In practice we are interested in representing the measurement in Cartesian coordinates, thus, the measurement is transformed again from the spherical coordinates to Cartesian coordinates $\mathbf{u}' = (x', y', z')$ as follows: $q = [(\hat{x}', \hat{y}', \hat{z}'), \Lambda_{xyz}]$, where

$$\hat{x}' = \hat{r}\sin\hat{\phi}\cos\hat{\theta} \quad ; \quad \hat{y}' = \hat{r}\sin\hat{\phi}\sin\hat{\theta} \quad ; \quad \text{and} \quad \hat{z}' = \hat{r}\cos\hat{\phi}$$

and the covariance matrix is:

$$\Lambda_{xyz} = \left(\frac{\partial(x, y, z)}{\partial(r, \phi, \theta)}\right) \Lambda_{r\phi\theta} \left(\frac{\partial(x, y, z)}{\partial(r, \phi, \theta)}\right)^t \quad .$$

The derivatives are taken at the point $(\hat{r}, \hat{\phi}, \hat{\theta})$. Here too, all values are known except for $\hat{r}$. Since the solution to the location problem incrementally improves the estimation of T, i.e. at step k there exists an estimate $\hat{T}_{k-1}$ from the previous step. We use this estimate to calculate an estimate of $\hat{r}$ at step k as follows:

$$\hat{r}_k = \|\hat{T}_{k-1}(\mathbf{u})\| \quad ,$$

where $\mathbf{u} \in M$ is the location of the corresponding point in the model. We emphasize that the uncertainty of this estimate, as expressed in the covariance matrix, is infinite.

The uncertainty generated from the raw input-data is propagated into uncertainty of the solution (i.e. estimated $3D$ location). Solution uncertainty denotes the belief we associate with the estimated location of the object. This uncertainty will be represented by a covariance matrix whose dimension is equal to the degrees of freedom of the transformation (6 for a rigid 3-D transformation). The optimal estimate will be that which minimizes the covariance matrix (or rather minimizes its norm). When representing the transformation as a 6 component vector, the dependence between the estimated rotation and estimated translation is expressed through the entries in the covariance matrix. This dependence is not considered in methods where the process of determining the rotation is separated from the process of determining the translation (e.g. [2], [12]). Taking this dependence into account is shown to give a more accurate result [13].

A. The System Definition

The variables to be estimated:
The representation of the transformation is composed of two components:

- The translation component is expressed by the vector $\mathbf{t}$;

$$\mathbf{t} = (t_x, t_y, t_z)^t \quad .$$

- The rotation component is described by a unit quaternion $\tilde{\mathbf{q}}$ [14]:

$$\tilde{\mathbf{q}} = (q_0, \mathbf{q}) = (q_0, q_1 i + q_2 j + q_3 k) \quad .$$

The rotation quaternion should satisfy the normality constraints: $\tilde{\mathbf{q}}\tilde{\mathbf{q}}^* = q_0^2 + \|\mathbf{q}\|^2 = 1$, where $\tilde{\mathbf{q}}^*$ is the conjugate of $\tilde{\mathbf{q}}$.

In practice we represent the rotation component by the vector: $\mathbf{s} \equiv \mathbf{q}/q_0$ from which the quaternion $\tilde{\mathbf{q}}$ can be reconstructed:

$$q_0 = \frac{1}{\sqrt{1 + \mathbf{s}^t\mathbf{s}}} \quad ; \quad \tilde{\mathbf{q}} = (q_0, q_0\mathbf{s}) \quad .$$

The vector $\mathbf{s}$ is a convenient representation of the rotational component; in addition to being minimal (having 3 parameters) the rotation equation is linear in $\mathbf{s}$ as will be shown later. In order to avoid singularities in the representation when $q_0 = 0$ we simultaneously use two different coordinate systems.

Considering these two components, the parameter vector to be estimated during the filtering process is:

$$\mathbf{T} = \left(\mathbf{s}^t, \mathbf{t}^t\right)^t \quad .$$

The observations:
A model point is represented by a vector $\mathbf{u}_i = (x, y, z)^t$ in an object centered frame of reference, where the index i denotes the step of the process at which this feature is considered (the same model point can be considered many times when there are several measurements of this point).
$\mathbf{u}'_i = (x', y', z')^t$ - is the real position of the point $\mathbf{u}_i$ in the viewer centered frame of reference.
$\hat{\mathbf{u}}'_i$ - is the measured position of the point $\mathbf{u}'_i$. This measurement is imprecise and can be represented as:

$$\hat{\mathbf{u}}'_i = \mathbf{u}'_i + \epsilon_i \quad ,$$

where ϵ_i is white noise satisfying:

$$E\{\epsilon_i\} = 0 \quad ; \quad E\{\epsilon_i \epsilon_i^t\} = \Lambda_i \quad ; \quad E\{\epsilon_i \epsilon_j^t\} = 0 \quad \forall i \neq j \quad .$$

The measurement model:
A mathematical relationship between the measured vector and the estimated vector is expressed, for each feature i, by a non linear quaternion equation:

$$\tilde{\mathbf{u}}'_i = \tilde{\mathbf{q}} \tilde{\mathbf{u}}_i \tilde{\mathbf{q}}^* + \tilde{\mathbf{t}} \quad , \tag{1}$$

where $\tilde{\mathbf{u}}_i, \tilde{\mathbf{u}}'_i, \tilde{\mathbf{t}}$ are quaternions associated with the vectors $\mathbf{u}_i, \mathbf{u}'_i, \mathbf{t}$ respectively. Given that $\tilde{\mathbf{q}}\tilde{\mathbf{q}}^* = 1$, multiplying Equation (1) by $\tilde{\mathbf{q}}$ yields:

$$\tilde{\mathbf{u}}'_i \tilde{\mathbf{q}} = \tilde{\mathbf{q}} \tilde{\mathbf{u}}_i + \tilde{\mathbf{t}} \tilde{\mathbf{q}} \quad .$$

Isolating the vector component of this quaternion equation and dividing by q_0 we get the matrix equation:

$$h_i(\mathbf{u}_i, \mathbf{u}'_i, \mathbf{T}) \equiv <\mathbf{u}'_i + \mathbf{u}_i> \mathbf{s} + (\mathbf{u}'_i - \mathbf{u}_i) - (I_3 + <\mathbf{s}>)\mathbf{t} = \mathbf{0} \quad , \tag{2}$$

where $\mathbf{s} \equiv \frac{\mathbf{q}}{q_0}$ as previously defined, I_3 is the 3×3 identity matrix and $<\cdot>$ denotes the matrix form of a cross product, i.e:

$$<\mathbf{v}> = \begin{pmatrix} 0 & -v_z & v_y \\ v_z & 0 & -v_x \\ -v_y & v_x & 0 \end{pmatrix} \quad ; \quad \textbf{and} \quad <\mathbf{v}>\mathbf{u} = \mathbf{v} \times \mathbf{u} \quad .$$

Notice that according to the definition of the measurement noise, we assume no correlation between the different measurement noise ($\text{cov}\{\epsilon_i, \epsilon_j\} = 0 \ \forall i \neq j$). This assumption is not always valid. When there is correlation between several measurements, we may consider these measurements as a single measurement by grouping the measurement values into a single vector and by combining their corresponding equations into a single vector equation.

B. The Estimation Control

The estimation process is composed of an incremental process, for which at each step $k-1$, there exists an estimate $\hat{\mathbf{T}}_{k-1} = \begin{pmatrix} \hat{\mathbf{s}}_{k-1} \\ \hat{\mathbf{t}}_{k-1} \end{pmatrix}$ of the transformation $\mathbf{T}$ and a covariance matrix Σ_{k-1} which represents the "quality" of the estimate $\hat{\mathbf{T}}_{k-1}$. Given a new match $(\mathbf{u}_k, \hat{\mathbf{u}}'_k)$ the current estimate is updated to be $\hat{\mathbf{T}}_k$ with the associated uncertainty Σ_k. The accuracy of the estimate increases, as additional matches are fused, i.e. $\Sigma_k \leq \Sigma_{k-1}$ ($\Sigma_{k-1} - \Sigma_k$ is nonnegative definite). The process terminates as soon as the uncertainty satisfies our criterion for accuracy or no additional match can be supplied [15].

Fusing the information from a match with the old estimate is performed using the *extended Kalman filter* (E.K.F.) process [16], [17].

Implementing the K.F. yields an unbiased estimate of $\mathbf{T}$ which is optimal in the linear minimal variance criterion [16]. In the case where the measurement noise ε is a Gaussian process (which is a reasonable assumption, considering the numerous sources of noise), the K.F. gives an estimate which is also the maximum-likelihood.

The convergence of the estimate to the true solution can be evaluated by studying the qualitative behavior of the covariance matrix Σ. The evolution of Σ during the process is given by (see [16]):

$$\Sigma_{t+1}^{-1} = H_{t+1}^t \Lambda_{t+1}^{-1} H_{t+1} + \Sigma_t^{-1} \quad .$$

Under the assumption that H and Λ are constant along time, we have:

$$\frac{\partial \Sigma^{-1}(t)}{\partial t} \approx \Sigma_{t+1}^{-1} - \Sigma_t^{-1} = H^t \Lambda^{-1} H$$

$$\text{so that:} \quad \Sigma(t) \sim \frac{(H^t \Lambda^{-1} H)^{-1}}{t}$$

i.e. the convergence is at a rate of t^{-1} and thus the squared deviation $\|\mathbf{T} - \hat{\mathbf{T}}^t\|^2$ also converges as t^{-1}.

A. Matching Control

The use of the K.F. process enables us to obtain reasonable matches during the estimation process [6], [18]; at each step i, use the current estimate $\hat{\mathbf{T}}_i$ and its corresponding confidence Σ_i to select "good" matches. The selection is done using goodness of fit tests:
Given a model-feature $\mathbf{u}_k$, let a candidate for a match be the measurement $(\hat{\mathbf{u}}'_k, \Lambda_k)$. According to this hypothesis, $h_k(\mathbf{u}_k, \hat{\mathbf{u}}'_k, \hat{\mathbf{T}}_{k-1})$ (Equation 2) is an independent random variable with a normal distribution which satisfies:

$$E\{h_k(\mathbf{u}_k, \hat{\mathbf{u}}'_k, \hat{\mathbf{T}}_{k-1})\} = E\{\hat{h}_k\} = 0$$

$$\text{Var}\{(\hat{h}_k)\} = (\frac{\partial h_k}{\partial \mathbf{u}'_k}) \Lambda_k (\frac{\partial h_k}{\partial \mathbf{u}'_k})^t + (\frac{\partial h_k}{\partial \mathbf{T}}) \Sigma_k (\frac{\partial h_k}{\partial \mathbf{T}})^t \equiv S_k \quad .$$

The "goodness" of fit between $\mathbf{u}_k$ and $\hat{\mathbf{u}}'_k$ is then given by the Mahalanobis distance:

$$d(\mathbf{u}_k, \hat{\mathbf{u}}'_k) = (\hat{h}_k) S_k^{-1} (\hat{h}_k)^t = g(\hat{\mathbf{T}}_k, \Sigma_k) \quad ,$$

where g has χ^2 distribution with $rank(S_k)$ degrees of freedom. The probability that the match is correct is inverse to g. As the process proceeds, the uncertainty Σ decreases lowering the number of acceptable matches.

B. Parallelization of the Process

Assume that the E.K.F. process was performed on two separate channels a and b, using n matches in each channel:

$$\{\hat{\mathbf{u}}'_i\}_{i=1\ldots n} \xrightarrow{E.K.F.} \hat{\mathbf{T}}_a, \Sigma_a$$

$$\{\hat{\mathbf{u}}'_i\}_{i=n+1\ldots 2n} \xrightarrow{E.K.F.} \hat{\mathbf{T}}_b, \Sigma_b$$

Optimal fusion of the $2n$ matches can be performed easily using the K.F. equations if we interpret $(\hat{\mathbf{T}}_a, \Sigma_a)$ as an *a-priori* estimation of $\mathbf{T}$ and consider $\hat{\mathbf{T}}_b$ as a "new measurement" with associated covariance matrix Σ_b. Extension of this method can fuse estimates obtained from a greater number of channels (requiring $\log m$ steps for m channels). It is thus possible to decompose the K.F. process into several channels and then fuse the obtained estimates into a single optimal solution. This framework can also be used to fuse information from several kinds of measured primitives, e.g. lines and planes, where each channel is dedicated to one kind of primitive. Simulation results (that can be seen in [19]) show that the parallel process converges to the same solution as the serial process.

We tested our method by simulating a model as a collection of points. The points of the model were chosen randomly from the $[0.100]^3$. The model points were transformed by a transformation $\mathbf{T}$ composed of a rotation s and a translation $\mathbf{t}$ limited in length to 200. The measurements of the transformed points were contaminated by white Gaussian noise. The algorithm estimates the transformation $\mathbf{T}$ using three types of measurements:

- $2D$ measurements from a perspective projection.
- $2D$ measurements from an orthographic projection.
- $3D$ measurements.

The algorithm assumes the correspondence between the points of the model and the measured points.

Typical results obtained by the simulations are presented in graphs 3-4. These graphs represent examples of measurement fusion from a single type and examples of measurement fusion from a mixture of types. When one type of measurement was used, a single measurement was fused at each step. In the case of mixed data types, perspective, orthographic and $3D$ measurements were fused together at each step. In the depicted simulations the s.t.d for perspective, orthographic and $3D$ measurements were 0.3, 7.0, and 8.0 respectively. In the case of perspective projection, noise was added to the measurements in the image plane. Gaussian noise in the image plane with s.t.d of 0.3 is equal in our camera configuration to 6% of the total size of the body as observed in the image plane. This corresponds to about 10 pixels of error in an object such as in Figure 2. The s.t.d. of the measurements were chosen such that $trace(\Sigma)$ as a function of the number of fused measurements is identical for the three simulations involving single data types (see graph 4-left in which the plots of trace Σ coincide). Thus, the contribution of each measurement to the quality of the pose estimate is equivalent for each type of data allowing a comparative evaluation of the fusion process.

Graphs 3-left and 3-right show the convergence of the estimates of the rotation $\hat{\mathbf{s}}$ and the translation $\hat{\mathbf{t}}$ as a function of the number of measurements (matched points). The vertical ordinate represents the normalized error of the estimate: $t_i^{error} = \frac{\|\hat{\mathbf{t}}_i - \mathbf{t}\|}{\|\mathbf{t}\|}$ in Graph 3-left and $s_i^{error} = \frac{\|\hat{\mathbf{s}}_i - \mathbf{s}\|}{\|\mathbf{s}\|}$ in Graph 3-right . Each of these graphs show the convergence in 4 cases: three cases depict fusion of a single data type and the fourth case depicts the convergence when the three types of measurements were fused together at each step. It can be seen that the convergence rate in the integrated case is much better than in the other cases. Graphs 4-left and 4-right depict the trace of the covariance matrix corresponding to the estimate $\hat{\mathbf{T}}$: $E\{\|\mathbf{T} - \hat{\mathbf{T}}\|^2\}$ in comparison with the squared deviation of $\hat{\mathbf{T}}$ from the true transformation: $\|\mathbf{T} - \hat{\mathbf{T}}\|^2$ (in the orthographic case the trace and the deviation are calculated without the last component of $\mathbf{T}$ which is the z component of the translation). The trace and the deviation shown in these graphs were averaged over 100 processes of 100 randomly generated objects. Graph 4-left shows the identical convergence of the traces for the three types of data which are fused separately, and the improvement in the trace in the integrated case. Graphs 4-left and 4-right show a high correlation between the estimate quality and the certainty its covariance matrix describes. This behavior indicates that the algorithm indeed exploits the supplied information. Moreover, it can be seen that the convergence rate of the estimate corresponds to a decay rate of t^{-1} as expected (see Section IV). Additional simulation results on various cases including simulations of the parallel process can be founded in [19].

Our algorithm was applied to measurements taken from $2D$ images of a battery charger (see Figure 2). The object model consists of 35 model points where the the location of each model point was measured relative to the front-bottom-left corner (the 3 edges are considered as the object centered coordinates). In the following example we took images of the object at four different positions. In all images the object was placed on a planar table (see the 4 pictures in Figure 2) and the real transformation between every two positions was measured (i.e. translation distance and angle of rotation). The algorithm was applied to each of the given images. The measurements consisted of 35 feature points; 29 features were $2D$ measurements taken from the image coordinate and 6 features were $3D$ measurements calculated by stereo triangulation. The 2D measurement noise was assumed to be a bivariate Gaussian process. The uncertainty of an image point measurement can be calculated by fitting a bivariate Gaussian to the local auto-correlation function of the point image [20]. The 3D measurement uncertainty can be easily derived from the image point uncertainty (for details see [21]). According to the results, the relative transformations between every two positions were calculated. The comparison between the real transformation and the constructed transformation as estimated by the algorithm is given in the following table. As can be seen, the results obtained by our algorithm are close to the real solution with a deviation of up to 0.9 cm in translation and 1.1° in rotation.

True Solution

	pose B	pose C	pose D
pose A	31.0 cm ; 32.5°	48.6 cm ; 24.0°	26.3 cm ; 46.0°
pose B		20.3 cm ; −8.5°	15.8 cm ; 13.5°
pose C			35.6 cm ; 22.0°

Estimated Solution

	pose B	pose C	pose D
pose A	30.2 cm ; 32.7°	47.7 cm ; 23.1°	26.8 cm ; 45.5°
pose B		20.5 cm ; −9.6°	16.4 cm ; 12.7°
pose C			35.4 cm ; 21.1°

In this paper we presented a new approach to estimating the pose of a rigid object in space, where no limitations are imposed on the dimensionality of the measurements and on the type of projection. The main advantages of the suggested approach are as follows:

- A uniform formulation for all types of measurements allowing simple and efficient fusion of information obtained from different types of sensors.
- Considering the spatial uncertainty of each measurement, in an explicit manner, enabling optimal exploitation of the available information from the measurements.
- The process additionally supplies an estimation of the quality of the solution. This quality estimate can assist in determining the number of measurements required for estimating the pose at a given precision.
- Fusing the measurements in an incremental process, thus easily incorporated into the matching process which is performed by a pruning-search of the interpretation tree. Additionally, the quality of the matching can be estimated by using statistical tests.

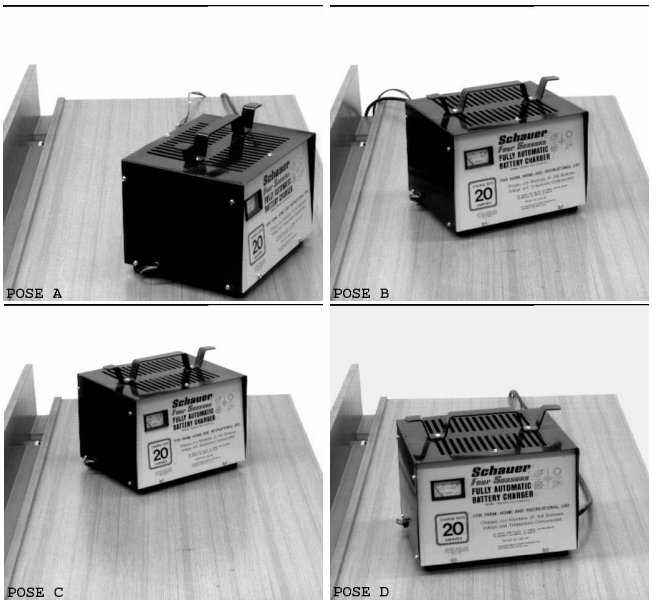

Fig. 2. Four images of four different positions of an object that were analyzed by our method. Top-left: pose A, top-right: pose B, bottom-left: pose C and bottom-right: pose D.

- The process can be easily parallelized.

Simulations of the described pose estimation process, showed quick and stable convergence of the estimate to the true solution. Good and stable solutions were also obtained for real models and images. We plan in the future to extend the results to more complicated features such as line and curved primitives. It is also possible to associate uncertainty with model-features due to imprecise modeling of the $3D$ object (for example, when modeling faces or other semi-elastic objects). In this paper we assumed an exact model with no uncertainties, but it is straightforward to include uncertainties in the model.

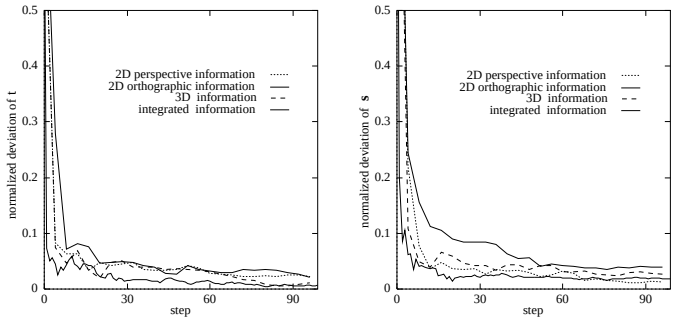

Fig. 3. Left graph: convergence of the normalized deviation of the translation estimate $\hat{\mathbf{t}}$. Right graph: convergence of the normalized deviation of the rotation estimate $\hat{\mathbf{s}}$. The graphs show 4 cases: fusing only perspective data, fusing only orthographic data, fusing only $3D$ data, and fusing all the above data types together.

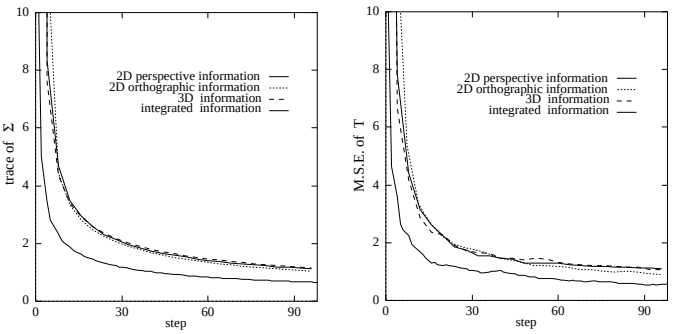

Fig. 4. Comparison between the trace of Σ (left graph) and the squared deviation of the estimate $\hat{\mathbf{T}}$ from the real transformation (right graph). The comparison is done for the 4 cases as in the graphs 4-left and 4-right. The trace and the deviation values were averaged over 100 simulations of 100 randomly generated objects.

[1] B. Sabata and J.K. Aggarwal, "Estimation of motion from a pair of range images: A review", *Computer Vision, Graphics and Image Processing*, vol. 54, no. 3, pp. 309–324, Nov 1991.

[2] R.M. Haralick, H. Joo, C.N. Lee, X. Zhuang, V.G. Vaidya, and M.B. Kim, "Pose estimation from corresponding point data", *IEEE Trans. Systems, Man, and Cybernetics*, vol. 19, no. 6, pp. 1426–1445, Nov/Dec 1989.

[3] R.Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses", *IEEE Journal of Robotics and Automation*, vol. 3, no. 4, pp. 323–344, Aug 1987.

[4] J.S.C. Yuan, "A general photogrammetric method for determining object position and orientation", *IEEE Transactions on Robotics and Automation*, vol. 5, no. 2, pp. 129–142, 1989.

[5] M.A. Fischler and R.C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography", *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, June 1981.

[6] O.D. Faugeras and M. Hebert, "The representation, recognition, and positioning of 3D shapes from range data", in *Techniques for 3D Machine Perception*, A. Rosenfeld, Ed., pp. 13–51. Elsevier Science, 1986.

[7] B.K.P Horn, "Closed-form solution of absolute orientation using unit quaternion", *J. Opt. Soc. Am.*, vol. 4, no. 4, pp. 629–642, 1987.

[8] K.S. Arun, T.S. Huang, and S.D. Blostein, "Least squares fitting of two 3D point sets", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 9, no. 5, pp. 698–700, Sept. 1987.

[9] S.D. Blostein and T.S. Huang, "Estimating 3-D motion from range data", in *Conf. on Artificial Intelligence Applications*, 1984, pp. 246–250.

[10] Z.C. Lin, T.S. Huang, S.D. Blostein, H. Lee, and E.A. Margerum, "Motion estimation from 3-D point sets with and without correspondences", in *Conference on Computer Vision and Pattern Recognition*, 1986, pp. 194–201.

[11] Y. Liu, T.S. Huang, and O.D. Faugeras, "Determination of camera location from 2D to 3D line and point correspondences", in *Conference on Computer Vision and Pattern Recognition*, 1988, pp. 82–88.

[12] W.E.L. Grimson and T. Lozano-Perez, "Model-based recognition and localization from sparse range or tactile data", *International Journal of Robotics Research*, vol. 3, no. 3, pp. 3–35, 1984.

[13] R. Kumar, *Model Dependent Inference of 3D Information from a Sequence of 2D Images*, PhD thesis, University of Massachusetts at Amherst, Feb. 1992.

[14] B.K.P. Horn, *Robot Vision*, MIT Press, 1986.

[15] Y. Hel-Or, A. Shmuel, and M. Werman, "Active feature localization", in *Active Perception and Robot Vision*. Springer Verlag, 1991.

[16] A.H. Jazwinski, *Stochastic Process and Filtering Theory*, Academic Press, 1970.

[17] P.S. Maybeck, *Stochastic Models, Estimation, and Control*, vol. 1, Academic Press, 1979.

[18] O.D. Faugeras, N. Ayache, and B. Faverjon, "A geometric matcher for recognizing and positioning 3D rigid objects", in *Conference on Artificial Intelligence Applications*, 1984, pp. 218–224.

[19] Y. Hel-Or, *Pose Estimation from Uncertain Sensory Data*, PhD thesis, Inst. of Computer Sciense, The Hebrew University of Jerusalem, 1993.

[20] A. Shmuel and M. Werman, "Active vision: 3D depth from an image sequence", in *International Conference on Pattern Recognition*, 1990, pp. 48–54.

[21] L. Matthies and T. Kanade, "The cycle of uncertainty and constraint in robot perception", *International Journal of Robotics Research*, vol. 4, 1987.