An Evidential Reasoning Framework for Object Tracking

Fabio Cuzzolin*, Ruggero Frezza

NAVLAB - Autonomous Navigation and Computer Vision Laboratory

Dipartimento di Elettronica ed Informatica, University of Padova, Italy

ABSTRACT

Object tracking consists of reconstructing the configuration of an articulated body from a sequence of images provided by one or more cameras. In this paper we present a general method for pose estimation based on the evidential reasoning. The proposed framework integrates different levels of description of the object to improve robustness and precision, overcoming the limitations of approaches using single-feature representations. Several image descriptions extracted from a single-camera view are fused together using the Dempster-Shafer "theory of evidence".[14] Feature data are expressed as belief functions over the set of their possibile values. There is no need of any a-priori assumptions about the model of the object. Learned refinement maps between feature spaces and the parameter space Q describing the configuration of the object characterize the relationships among distinct representations of the pose and play the role of the model. During training the object follows a sample trajectory in Q. Each feature space is reduced to a discrete frame of discernment (FOD) and refinements are built by mapping these FODs into subsets of the sample trajectory. During tracking new sensor data are converted to belief functions which are projected and combined in the approximate state space. Resulting degrees of belief indicate the best pose estimate at the current time step. The choice of a sufficiently dense (in a topological sense) sample trajectory is a critical problem. Experimental results concerning a simple tracking system are shown.

Keywords: *articulated object tracking, belief function, refinement, sample trajectory, learning.*

1. INTRODUCTION

Object tracking is an interesting field of computer vision whose difficult problems stimulate the search for new viewpoints and suitable mathematical tools. It consists on reconstructing the actual pose a moving object by processing the sequence of images taken during the movement.

Several different approaches to this task has been developed: model-based tracking algorithms[12] exploit optimization techniques in order to minimize a residual between predicted and real measurements and achieve the set of parameters that better fits the new image evidences. They often suffer the problematic convergence of numerical minimization algorithms and generally need to be manually initialized. Other techniques[1] aim to give qualitative descriptions of motions in order to receive messages or recognize meanings; gesture recognition[2] is a typical example.

All these methods are generally based on a single kind of feature which can produce misunderstanding under particular conditions (rapid movements, for instance). Besides, model-based feature extraction can hardly carry independent evidence about the image for it is driven by current parameter estimates.

What we are looking for is a tracking system which rests on information about images as complete as possible. It should integrate different descriptions to increase the estimation robustness and overcome single-feature drawbacks. It should also measure the consistency of the acquired data and compute the estimate from the most coherent set of measurements. It must be pointed out that it is often impossible to write analytic relations between different features for they can concern completely unrelated aspects of the images.

The *theory of evidence*[14] has been introduced in the late Seventies by Glenn Shafer as a way of representing epistemic knowledge, starting from the seminal work[3] of Arthur Dempster. In this formalism the best representation of chance is a *belief function* (b.f.) rather than a Bayesian mass distribution. They assign probability values to *sets* of possibilities rather than single events: their appeal rests on the fact they naturally encode evidence in favor to *propositions*.

The theory provides a simple method for combining the evidence carried by a number of different sources (*Dempster's rule*) with no need of any *a-priori* distributions. A formal definition of the different levels of detail in knowledge representation is introduced, when the concept of *family of compatible frames* reflects the intuitive idea of different

*Email: cuzzolin@dei.unipd.it; Phone: +39-049-8277834

descriptions (features) of a same phenomenon.
In Section 2 we will introduce the evidential reasoning notions we consider significant for our purposes. In Section 3 the theoretical foundations of our tracking framework will be shown. In particular we will describe the way ideas like relationships between feature spaces and parameter spaces take place in this formalism and some methods for expressing a measurement as a belief function. Section 4 will concern the measure of consistency among feature data and the way the corresponding belief functions are combined. We will see a method for extracting a pointwise estimate of the body configuration from a b.f. and show how to build an *evidential model* of the object to track its pose. In Section 6 experimental results concerning a simple planar robot will be illustrated. Finally some further developments of our evidential approach will be discussed.

2. THE THEORY OF EVIDENCE

2.1. Notion of Belief

Following Shafer[14] we will call the finite set of possibilities *frame of discernment* (FOD).

DEFINITION 2.1. *A* basic probability assignment *over a FOD* Θ *is a function* $m : 2^\Theta \to [0,1]$ *such that*

$$m(\emptyset) = 0,\ \sum_{A \subset \Theta} m(A) = 1.$$

The elements of 2^Θ associated to non-zero values of m are called *focal elements* and their union *core*. Now suppose a b.p.a. is introduced over an arbitrary FOD.

DEFINITION 2.2. *The* belief function *given by the basic probability assignment* m *is defined as:*

$$Bel(A) = \sum_{B \subset A} m(B)$$

In the theory of evidence a probability function is simply a peculiar belief function which satisfies the additivity rule for disjoint sets. It can be proved that a function Bel is Bayesian $\Leftrightarrow \exists\ p : \Theta \to [0,1]$ such that

$$\sum_{\theta \in \Theta} p(\theta) = 1,\ Bel(A) = \sum_{\theta \in A} p(\theta)$$

for all $A \subset \Theta$. Belief functions representing distinct bodies of evidence are combined together by means of the *Dempster's rule of combination*:

the orthogonal sum $Bel_1 \oplus Bel_2$ *of two belief functions is a function whose focal elements are all the possible intersections between the combining focal elements and whose b.p.a. is given by*

$$m(A) = \frac{\sum_{i,j:A_i \cap B_j = A} m_1(A_i) m_2(B_j)}{1 - \sum_{i,j:A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j)}$$

If we examine the normalization constant in the above expression we discover it has an intuitive meaning. In fact, it measures the *level of conflict* between belief functions for it represents the amount of probability they attribute to contradictory (i.e. disjoint) subsets.

DEFINITION 2.3. *We call* level of conflict *between* Bel_1 *and* Bel_2 *the logarithm of the normalizing constant of Demspter's rule*

$$\mathcal{K} = log \frac{1}{1 - \sum_{i,j:A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j)}$$

The above concepts are easily extended to the general case of combining several belief functions.

2.2. Families of FOD and refinings

One of the amazing ideas in the D-S theory is the simple claim that our knowledge is inherently imprecise. New amount of evidence could allow us to make deductions over more refined environments.

This argument is embodied into the notion of *refinement*.

DEFINITION 2.4. *Given two frames* Θ *and* Ω*, a map* $\omega : 2^\Theta \to 2^\Omega$ *is a* refining *if it satisfies the following conditions:*

$$\omega(\{\theta\}) \neq \emptyset \ \forall \theta \in \Theta, \quad \omega(\{\theta\}) \cap \omega(\{\theta'\}) = \emptyset \ if \ \theta \neq \theta', \quad \cup_{\theta \in \Theta} \omega(\{\theta\}) = \Omega.$$

The finer frame is called a *refinement* of the first one and we call Θ a *coarsening* of Ω. Two maps are associated to a given refining: the *inner reduction* $\underline{\theta}(A) = \{\theta \in \Theta | \omega(\{\theta\}) \subset A\}$ gives the largest subset of Θ implying a proposition A when the *outer reduction* $\bar{\theta}(A) = \{\theta \in \Theta | \omega(\{\theta\}) \cap A \neq \emptyset\}$ is the smallest subset of Θ that is implied by A. These applications play an important role in our feature integration method that has its center in the notion of *families of compatible frames*.

DEFINITION 2.5. *A non-empty collection of finite non-empty sets* $\mathcal{F}$ *is a* family of compatible frames of discerment *with refinings* $\mathcal{R}$*, where* $\mathcal{R}$ *is a non-empty collection of refinings between couples of frames in* $\mathcal{F}$ *if* $\mathcal{F}$ *and* $\mathcal{R}$ *satisfy the following requirements:*

1. composition of refinings*: if* $\omega_1 : 2^{\Theta_1} \to 2^{\Theta_2}$ *and* $\omega_2 : 2^{\Theta_2} \to 2^{\Theta_3}$ *are in* $\mathcal{R}$*, then* $\omega_1 \circ \omega_2$ *is in* $\mathcal{R}$*;*
2. identity of coarsenings*: if* $\omega_1 : 2^{\Theta_1} \to 2^\Omega$ *and* $\omega_2 : 2^{\Theta_2} \to 2^\Omega$ *are in* $\mathcal{R}$ *and* $\forall\ \theta_1 \in \Theta_1, \exists\ \theta_2 \in \Theta_2$ *such that* $\omega_1(\{\theta_1\}) = \omega_2(\{\theta_2\})$ *then* $\Theta_1 = \Theta_2$ *and* $\omega_1 = \omega_2$*;*
3. identity of refinings*: if* $\omega_1 : 2^\Theta \to 2^\Omega$ *and* $\omega_2 : 2^\Theta \to 2^\Omega$ *are in* $\mathcal{R}$*, then* $\omega_1 = \omega_2$*;*
4. existence of coarsenings*: if* $\Omega \in \mathcal{F}$ *and* $A_1, ..., A_n$ *is a disjoint partition of* Ω *then there is coarsening in* $\mathcal{F}$ *which corresponds to this partition;*
5. existence of refinings*: if* $\theta \in \Theta \in \mathcal{F}$ *and* $n \in \mathcal{N}$ *then there exists a refining* $\omega : 2^\Theta \to 2^\Omega$ *in* $\mathcal{R}$ *such that* $\omega(\{\theta\})$ *has* n *elements;*
6. existence of common refinements*: every pair of elements in* $\mathcal{F}$ *has a common refinement in* $\mathcal{F}$*.*

A collection of compatible frames has many common refinements, but one of these is particularly simple. This unique FOD is called the *minimal refinement* of the collection $\Theta_1, ..., \Theta_n$ and it is the simplest space where you can compare propositions represented by subsets of two different compatible frames.
A given body of evidence determines a belief function over *every* frame of the family, i.e. it gives a *family of compatible functions* but the combination of evidence can produce quite different results depending where it is computed.

DEFINITION 2.6. *If* S_1 *and* S_2 *are support functions over a frame* Ω *and* $\bar{\theta} : 2^\Omega \to 2^\Theta$ *is an outer reduction,* Θ *is said to* discern the relevant interaction of S_1 and S_2 *if* $\bar{\theta}(A \cap B) = \bar{\theta}(A) \cap \bar{\theta}(B)$ *whenever* A *is a focal element of* S_1 *and* B *is a focal element of* S_2*.*

It can be proved that in these hypothesis $(S_1|2^\Theta) \oplus (S_2|2^\Theta) = (S_1 \oplus S_2)|2^\Theta$ i.e the functions can be combined over Θ with no loss of information.

3. THEORETICAL FRAMEWORK

Several concepts of the Demspter-Shafer theory are very attractive in the perspective of an *information fusion* approach to object tracking. The basis notion of family of belief functions nicely fits the idea of different *kinds of representation* of motion. Precision levels of description can be formally defined by refinements when refining maps play the role of relationships among distinct feature spaces. Dempster's rule provides a simple method for combining evidence and a measure of *conflict* of the acquired data is also provided. Unfortunately, the theory developed by Shafer in the late Seventies was restricted to the case of *finite* sets of possibilities. Several attempts[8] of extending the theory to continuous frames were made (see for example Kohlas' "theory of hints[9]" or the notion of *random set*[1076]) but they seem far to constitute an ultimate answer in the search for a general theory of evidence. Thus we are constrained to find a discrete framework[15] in which combinate generally continuous measurements.
This involves several tasks:

1. building a *discrete* frame of discernment approximating a continuous feature space;

2. transforming the acquired feature data into *belief functions* over the appropriate FOD;

3. measuring their *conflict level* and rejecting the discarding features;

4. *combining* these functions in a common environment;

5. extracting an *object configuration estimate* from the resulting b.f.

The choice of a method for feature discretization[11] is of course *critical*: naive approaches like, for example, the regular partitioning of the measured feature range are clearly inadequate for they introduce arbitrary subdivisions of feature ranges and waste the information carried by the measurements.
We will see in the following how the well known formalism of the *hidden Markov models* fits our requirements.

3.1. Sample trajectory

As we have seen the evidential reasoning requires building discrete descriptions of the spaces involved by the problem. In particular we need to approximate the parameter space $\mathcal{Q}$ of the object, whose points represents internal configurations of the body. For example a manipulator with d degrees of freedom $\mathcal{Q}$ will have in general a domain of $\mathcal{R}^d$ as configuration space. Excluding pathological situations we can assume the *compactness* (the parameter space has no unreachable limit poses) and the *connectness* of $\mathcal{Q}$ (every shape can be assumed by an articulated object starting from any initial pose). The domain will not be in general linearly connected: it can have cuts or holes due to position constraints concerning two or more parts of the object. Even a kinematic model of the object is not sufficient to fully describe the configuration space, for these constraints are hard to formulate and depend on the size of body.
We can overcome these obstacles by describing a *sample trajectory* in the parameter space of the tracked object, i.e. collecting a significant set of configuration points

$$\tilde{\mathcal{Q}} \doteq \{q(t_k), k = 1, ..., T\}$$

for example by sampling a continuous curve $\gamma \subset \mathcal{Q}$. $\tilde{\mathcal{Q}}$ must satisfy two conditions:

1. it has to be *dense* in $\mathcal{Q}$:
$$\forall q \in \mathcal{Q} \; \exists \; k \; s.t. \; \|q - q(t_k)\| < \epsilon.$$

2. each feature function y_i cannot have sharp variations in the ϵ-neighbourhood of a sample point.

The first condition simply claims that if we chose the object estimated position in $\tilde{\mathcal{Q}}$ we commit an error not greater than ϵ. The second one ensures that each sample is a good representative of its neighbours for its feature value is similar to the others in the ϵ-neighbourhood. The following classical result demonstrates that such a trajectory exists, at least for non-pathological situations:

THEOREM 3.1. (**Small oscillations theorem**) *If a function $f : \mathcal{Q} \to \mathcal{R}$ is continuous over a compact subset $\mathcal{Q} \subset \mathcal{R}^n$ then*

$$\forall \epsilon > 0, \; \exists \, partition \; \Pi \; of \; \mathcal{Q} \mid if \; m_k = \min_{q \in \mathcal{Q}_k} f(q) \; and \; M_k = \max_{q \in \mathcal{Q}_k} f(q) \; \Rightarrow \; M_k - m_k < \epsilon \; \; \forall \mathcal{Q}_k \in \Pi.$$

It suffices to choose a representative point for each element of the partition $\Pi = \{\mathcal{Q}_k, \; k = 1, ..., T\}$. The thesis holds even if f has first order discontinuities.
Looking at Theorem 2.6 suggests a charming analogy. The above partition induces a refining ω_Q between the configuration space and the sample trajectory

$$\omega_Q : \tilde{\mathcal{Q}} \to 2^{\mathcal{Q}}, \;\; \omega_Q(\{q(t_k)\}) \doteq \mathcal{Q}_k$$

so we can define the analogous of the outer reduction

$$\bar{\theta}(X) \doteq \{q(t_k) \in \tilde{\mathcal{Q}} \mid \omega_Q(\{q(t_k)\}) \cap X \neq \emptyset\}.$$

THEOREM 3.2. $\tilde{\mathcal{Q}}$ discerns the relevant interactions *between a pair of feature functions* $y_1 : \mathcal{Q} \to \mathcal{Y}_1,\ y_2 : \mathcal{Q} \to \mathcal{Y}_2$ *defined over* $\mathcal{Q}$, *i.e.*

$$\bar{\theta}(y_1^{-1}(A) \cap y_2^{-1}(B)) = \bar{\theta}(y_1^{-1}(A)) \cap \bar{\theta}(y_2^{-1}(B)) \ \ \forall A \subset \mathcal{Y}_1, B \subset \mathcal{Y}_2.$$

if and only if

$$y(\tilde{\mathcal{Q}}_k) = y_1(\tilde{\mathcal{Q}}_k) \times y_2(\tilde{\mathcal{Q}}_k) \ \ \forall k$$

where

$$\begin{array}{rcccc} y & : & \tilde{\mathcal{Q}} & \to & \tilde{\mathcal{Y}}_1 \times \tilde{\mathcal{Y}}_2 \\ & & q & \mapsto & (y_1(q), y_2(q)). \end{array}$$

Of course this is not a condition on the sample trajectory but if condition 2. is satisfied with ϵ small enough the probability of wrong inferences on $\tilde{\mathcal{Q}}$ is reduced. This argument in a sense extends to continuous domains the evidential reasoning language and gives a precise characterization of the intuitive idea of sample trajectory. What is really important here is the relationship between parameter and feature spaces when the topology of $\mathcal{Q}$ gives only a constraint $\tilde{\mathcal{Q}}$ must satisfy.
We already know that we acnnot write the feature maps $f_i : \mathcal{Q} \to \mathcal{Y}_i$ analytically. This suggests a "learning" algorithm to make the sample trajectory more dense in the regions where features have high rates of change:

1. first a suitable value ϵ is chosen;
2. a sample trajectory satisfying condition 1. is followed;
3. the N feature collections $\{y_i(t),\ 0 \leq t \leq T\}$ (feature matrices) are computed;
4. the rate of change of each feature is calculated;
5. if the matrices satisfy condition 2. the algorithm terminates;
6. otherwise new samples in the high-variation zones are added and we come back to point 2.

3.2. Hidden Markov models

A *hidden Markov model*[4] (HMM) is a stochastic dynamical system whose sequence of state $\{X_k\}$ form a *Markov chain*; the only observable quantity is a corrupted version y_k of the state called *observation process*:

$$\left\{ \begin{array}{l} X_{k+1} = AX_k + V_k \\ y_{k+1} = CX_k + diag(W_k)\Sigma X_k \end{array} \right.$$

here $\{V_k\}$ and $\{W_k\}$ are sequences of i.i.d. gaussian noises.
A fundamental property of this class of models is the capability of self-learning the set of parameters A, C and Σ given a sequence of observations that are supposed to be produced by the system. This *expectation-maximization* algorithm is based on a iterative update of the matrices: at each loop the entire sequence of data is processed.
Once established the best parameter values the HMM output is the state estimate associated to the current observation. The estimate is obtained by measuring the probabilistic distance $\{\Gamma_j(y_k)\}$ of the measurement from each state representative CX_j in the observation space.

3.3. FOD Realization

The expectation-maximization algorithm provides a simple method for building the frame of discernment which better approximates a continuous feature space $\mathcal{Y}_i$. The features $\{y_k^i\}$ extracted from the images of the sample trajectory are collected together and passed as input to a Markov model with n_i states: after the learning procedure a set of n_i densities over the feature space is set up.[2] Hence these densities are equivalent to an *implicit* partition of the range: each feature point is attributed to the state with the highest value of Γ_j. This way the partition borders are automatically traced by the EM algorithm, following the clusters actually formed by the sample data.
The set of states $\{X_1, ..., X_{n_i}\}$ of the HMM finally forms the frame of discernement Θ_i approximating the i-th feature space:

$$\Theta_i = \{\theta_1, ..., \theta_{n_i}\} \leftrightarrow X = \{X_1, ..., X_{n_i}\} \leftrightarrow \pi_{Y_i} = \{Y_i^1, ..., Y_i^{n_i} | Y_i^j = \{y \in Y_i \ s.t. \ \Gamma_j(y) > \Gamma_l(y) \ \forall \ l \neq j\}\}$$

3.4. Building refinings

The other output of the EM algorithm is the sequence of state estimates associated to the sample sequence. Each feature vector $y_i(t_k)$ is "attributed" to a single state $\hat{X}_i(t_k) \in \{\theta_i^1, ..., \theta_i^n\}$ of the model.
Each element θ_i^j of the FOD (discretized feature) is then naturally associated to a set of sample points in the parameter space (or equivalently to a collection of time instants):

$$\theta_i^j \in \Theta_i \leftrightarrow \tilde{Q}_i^j \doteq \{t_k | \hat{X}_i(t_k) = \theta_i^j\}.$$

It is easy to see that the collection $\pi_i^Q = \{\tilde{Q}_i^j,\ j = 1, ..., n_i\}$ forms a partition of the sample trajectory: for every *measurement space* Θ_i the multi-valued map

$$\omega_i : \Theta_i \to \tilde{Q},\ \omega_i(\{\theta_i^j\}) = \tilde{Q}_i^j$$

is then a refining to the approximate state space $\tilde{Q}$.

3.5. System architecture

Now we can give a comprehensive description of our tracking system. The feature spaces and the collection $\tilde{Q}$ of points of the sample trajectory form a finite subset of a family of compatible frames of discernment. $\tilde{Q}$ is a common refinement of the collection of feature spaces but not the *minimal* refinement because of the presence of groups of undistinguished samples.
This family of frames along with their refining maps is characteristic of the articulated body: we call it the *evidential model* of the object.
The following scheme summarize our framework and the relationships among the involved spaces

$$\begin{array}{ccccc}
 & & \Pi^{Y_i} = \{\mathcal{Y}_i^1, ..., \mathcal{Y}_i^T\} & \overset{ref}{\longleftarrow} & \pi^{Y_i} = \{Y_i^1, ..., Y_i^{n_i}\} \\
 & & \updownarrow & & \updownarrow \\
\mathcal{Y}_i & \subset & \tilde{\mathcal{Y}}_i = \{y_i(q(0)), ..., y_i(q(T))\} & \overset{\hat{x}}{\longrightarrow} & X_i = \{x_1, ..., x_{n_i}\} \\
y_i \uparrow & & y_i \uparrow & & \updownarrow \\
\mathcal{Q} & \overset{\omega_Q}{\longleftarrow} & \tilde{\mathcal{Q}} = \{q(0), ..., q(T)\} & \overset{\omega_i}{\longleftarrow} & \Theta_i = \{\theta_1^i, ..., \theta_{n_i}^i\} \\
 & & \updownarrow & & \updownarrow \\
 & & \Pi^{\mathcal{Q}} = \{\mathcal{Q}_1, ..., \mathcal{Q}_T\} & \overset{ref}{\longleftarrow} & \pi_i^{\mathcal{Q}} = \{Q_i^1, ..., Q_i^{n_i}\}
\end{array}$$

where *ref* indicates refinings between partitions (not to be confused with refinings between *FODs*).

3.6. Building belief functions from measurements

The other step in our evidential approach to object tracking is the translation of the *continuous* measurements we get from image analysis into belief functions (called *measurement functions*) over the finite measurement spaces Θ_i we introduced to approximate the feature spaces:

$$\begin{array}{ccccc}
y_i & \longrightarrow & s_i : & \mathcal{P}(\Theta_i) & \to & [0,1] \\
 & & & A \subset \Theta_i & \mapsto & s_i(A)
\end{array}$$

We have seen how to train a HMM and use it to make an *implicit* discretization of a feature space $\mathcal{Y}_i$. Given a feature vector y_i^k as input it produces as output the set of *likelihoods* $\{\Gamma_j(y_i^k)\}_{j=1,...,n_i}$ of the measurement with respect to each of the n_i states of the model. We propose here two possible methods to build a measurement belief function from a set of likelihoods.

3.6.1. Bayesian constructor

The *Bayesian* constructor simply assigns to each element of the FOD the normalized value of the likelihood associated to the correspondent state of the HMM:

$$m_{s_i}(\{\theta_i^j\}) = \frac{\Gamma_j(y)}{\sum_{k=1}^n \Gamma_k(y)}.$$

3.6.2. Shafer's constructor

The second technique comes out by imposing these two conditions to the resulting belief function:

1. it has to preserve the *relative likelihoods* of the measurements;
2. it must belong to the class of consonant b.f., i.e. belief functions whose focal elements are nested: $\mathcal{A}_1 \subset ... \subset \mathcal{A}_f$.

Shafer has proved that there is *exactly one* belief function satisfying these hypothesis. If $\{\Gamma_j(y)\}_{j=1,...,n_i}$ are the likelihood values associated to an observation y then this unique b.f. is given by

$$s_i(A) = 1 - \frac{\max_{\theta_i^j \in \bar{A}} \Gamma_j(y)}{\max_{\theta_i^j \in \Theta_i} \Gamma_j(y)}$$

The consonance assumption ensures that a set of discretized measurements receives a high degree of support only if it includes a large number of elements θ with high likelihood.

4. COMBINING MEASUREMENT FUNCTIONS

Measurement functions defined over distinct feature spaces must be projected over their common refinement $\tilde{\mathcal{Q}}$ in order to drive the estimation.

DEFINITION 4.1. *Given a belief function s defined over a FOD Θ and a refining $\omega : 2^\Theta \to 2^\Omega$ to another frame Ω the function $\mathcal{S}$ over Ω is called* vacuous extension *of s into Ω if*

$$s(A) = \mathcal{S}(\omega(A)) \ \ \forall A \subset \Theta.$$

Once the measurement functions are projected onto the approximate parameter space the can be combined by using the Dempster's rule to fuse the information they carry. Unfortunately, there is no way to ensure such a comprehensive belief function exists.

THEOREM 4.2. *Let $\Theta_1, ..., \Theta_n$ be a set of compatible FODs. Then all the possible sets of belief functions $s_1, ..., s_n$ defined over $\Theta_1, ..., \Theta_n$ respectively are combinable over their minimal refinement $\Theta_1 \otimes ... \otimes \Theta_n$ iff $\Theta_1, ..., \Theta_n$ are* independent*, i.e. $\forall A_1 \subset \Theta_1, ..., A_n \subset \Theta_n \ \ \omega_1(A_1) \cap ... \cap \omega_n(A_n) \neq \emptyset$.*

Proof. $\Rightarrow$: since an arbitrary collection $\{s_i\}$ possesses an orthogonal sum we can choose a set of trivial b.f. focusing on a single subset A_i, so that $\omega_1(A_1) \cap ... \cap \omega_n(A_n) \neq \emptyset$ because of their combinability. Varying the focal elements $\{A_i\}$ arbitrarily we have the FODs must be independent.
$\Leftarrow$: if $\Theta_1, ..., \Theta_n$ are independent then

$$\forall A_1 \subset \Theta_1, ..., A_n \subset \Theta_n \ \ \omega_1(A_1) \cap ... \cap \omega_n(A_n) \neq \emptyset$$

so that whenever you choose a belief function on each frame and a focal element for each function their intersection could not be empty. □

It can be proved that if $\Theta_1, ..., \Theta_n$ are independent then $\Theta_1 \otimes ... \otimes \Theta_n = \Theta_1 \times ... \times \Theta_n$ so the above result shows that *the combination is guarantee only for trivially interacting feature spaces.* Hence at each time instant any arbitrary set of measurement functions will be characterized by a *level of conflict* in the range $[0, +\infty]$. If $\mathcal{K} = +\infty$ the corresponding belief functions will have no orthogonal sum.
We need a method for detecting the most coherent groups of b.f.'s and choosing one of them in order to calculate the object configuration estimate.

4.1. Conflict graph and sets of compatible features

A basis property of the level of conflict is[14]:

THEOREM 4.3.

$$\mathcal{K}(s_1, ..., s_{n+1}) = \mathcal{K}(s_1, ..., s_n) + \mathcal{K}(s_1 \oplus ... \oplus s_n, s_{n+1})$$

Thus if $\mathcal{K}_{s_i,s_j} = +\infty$ then $\mathcal{K}_{s_i,s_j,s_k} = +\infty \ \ \forall s_k$, i.e. if a pair of functions is not combinabile then all the sets of functions including this pair cannot produce any orthogonal sum. This suggests a *bottom-up* tecnique:

1. first the level of conflict is computed for each pair of measurement function $(s_i, s_j),\ i, j = 1, ..., n$;
2. then a *conflict graph* is build in the following way: once decided a suitable threshold level for the conflict every graph node represents a feature function when edges between them indicates a conflict level below the threshold;
3. the subsets of combinabile b.f. of size $d+1$ are *recorsively* computed from those of size d by checking whether the addition of any other node to the set generates a completely connected subgraph.

This algorithm produces a number of groups of b.f. over $\tilde{Q}$ that can be chosen to generate our estimate.

4.2. Feature group choice criteria

We have considered three different ways to select a candidate set of features:

1. *largest subset*: the group of features of the biggest size is selected;
2. *lowest conflict*: we choose the subset with the lowest level of conflict and size over a certain threshold;
3. *splitting*: all the alternative estimates generated by groups biggest enough are computed.

It should be noted that during the normal evolution of the system there is only one large set of coherent features even if a few measurements can deviate from the consensus. The presence of several candidate groups with similar size reveals a pathological situation.
The above feature selection methods can be tested by taking in account the associated estimate errors. Adopting the splitting strategy we can detect in the same way the *most reliable* set of features and give it a larger weight in the combination process.

4.3. Learning the conflict threshold

One may think that arbitrary choices of the conflict threshold can perhaps produce any desired estimated value. On the other hand, there is no theoretical argument supporting a precise assignment for it.
An interesting approach leaves this choice to a *learning* procedure:

- after the implementation of the family of FODs (feature spaces, approximate parameter space and refinings) a new trajectory is followed;
- at each time instant the measurement functions are combined (setting $\mathcal{K}= 0$ and summing all the compatible b.f.) and the estimates are produced;
- the estimated parameter sequence is compared with the actual parameter sequence and the desired threshold value is obtained by taking the mean value of the sequence of conflict levels weighted by the estimation errors $|q - \hat{q}|$.

The resulting conflict threshold will then be

$$\hat{\mathcal{K}} = \sum_{k=0}^{T} \mathcal{K}_{c(s_1,...s_n)}(k)|q(k) - \hat{q}(k)|$$

where $c(s_1, ..., s_n)$ indicates the subset of compatible measurement functions.

4.4. Pointwise estimation

Combining evidence over the common refinement $\tilde{\mathcal{Q}}$ produces a complex belief function as current object pose estimate. We would like to achieve a more intuitive *pointwise* value in order to compare it with the actual measured parameters.
Our approach consists on approximating the "epistemic" estimate with a Bayesian belief function and then calculating the mean parameter value simply from a sum of the samples associated to singletons in $\tilde{\mathcal{Q}}$ weighted by their probabilities:

$$\hat{q} = \sum_{k \in \tilde{Q}} p(k)q(t_k).$$

DEFINITION 4.4. *The plausibility function associated to a belief function s*

$$Pl_s(\{A\}) \doteq 1 - s(\Theta \setminus \{A\})$$

expresses the degree to which the evidence does not impugn a given proposition.

The above definition suggests to adopt the relative plausibilities of the singletons in $\tilde{\mathcal{Q}}$ to compute the pointwise estimate. Finally we have

$$\hat{q} = \sum_{k \in \mathcal{C}_s} Pl_s(q(t_k))q(t_k)$$

where $s = s_1 \oplus ... \oplus s_n$ is the belief estimate of the object pose.

5. ALGORITHMS

We can now summarize our object tracking algorithm.

5.1. Training

In the *training* phase the evidential model of the object is built. The topology of the sample trajectory is iteratively modified by increasing the number of samples belonging to regions where the features have high rates of change: this procedure has been fully described in Section 3.1.
Once the best trajectory is achieved we:

1. calculate the final feature matrices;
2. process every feature matrix by a HMM which produces:
   the model of the correspondent FOD Θ_i;
   the refinings ω_i between each FOD and the approximate parameter space $\tilde{Q}$.

5.2. Tuning

The evidential model produced by the training algorithm still depends on a number of parameters, i.e. the dimensions $\{n_i,\ i = 1, ..., n\}$ of the approximate feature spaces and the conflict threshold. The adequate number of state for each feature space can be calculated by analyzing the clusters the sample data form or alternatively learned from the measured estimation error as n_i increases.
Even if this is formally a multivariable optimization problem it seems reasonable to calculate each single cardinality $\hat{n}_i$ *separately* for it represents an inherent property of the feature. The level of conflict, instead, is estimated by means of the procedure exposed in Section 4.3.

5.3. Tracking

Given the evidential model of the moving body, tracking an arbitrary motion reduces to the following steps: for each time instant k

1. the n feature vectors $\{y_i^k\}_{i=1,...,n}$ are computed from the acquired image;
2. each feature vector y_i^k is processed by the corresponding HMM which produces a set of likelihood values $\{\Gamma_i^j\}$;
3. a measurement belief function s_i is built from these likelihoods;
4. all these functions $\{s_i\}$ are projected onto the discrete parameter space $\tilde{\mathcal{Q}}$;
5. the most coherent set of features is detected by analyzing their conflict graph and their orthogonal sum $s = s_1 \oplus ... \oplus s_n$ (*estimate function*) is calculated;
6. the pointwise estimate of the object configuration is obtained by means of Bayesian approximation.

In the following Section we will compare these estimates to actual configurations of a simple moving body.

6. EXPERIMENTAL RESULTS

Let us see the experimental behaviour of our information integration framework in a simple situation.
The articulated object we adopted is the 2 d.o.f. planar robot PantoMouse built by the Industrial Electronics group of our department. Its end-effector is a little metallic cylinder which can move inside a 19×19 mm area. We acquired 170×120 gray-level images with a Sony XC-75 CCD camera controlled by a National Instruments IMAQ PCI-1408 frame grabber.
We have taken pictures of the end-effector rectangular workspace only (see Figure 1a). The sample trajectory and a number of other robot motion were first acquired in batch mode by using the NI-IMAQ library attached to Labview when training and tracking routines were written in Matlab.

6.1. Features

We have chosen for image representation an interesting topology-based kind of feature called *size function*[5] which has been introduced by P. Frosini for robust representation of contours (see Figure 1). First[2] we extract the silhouette of

a) b) c) d)

Figure 1. The feature extraction process. a) Real image. b) Image of the edges. c) The measurement function associated to the selected line. d) The corresponding size function table.

the moving body and compute a collection of *measuring functions* called "distance from a line". They are obtained by simply projecting the contour over a sheaf of lines passing through its center.

DEFINITION 6.1. *Given a continuous function $\varphi : D \to \mathcal{R}$ the corresponding size function is a map $\mathcal{L}_\varphi : \mathcal{R} \times \mathcal{R} \to \mathcal{N}$ which, for each pair $(x, y) \in \mathcal{R} \times \mathcal{R}$ gives the number of connected components of the sub-domain $D_y \doteq \{d \in D : \varphi(v) \leq y\}$ having non-empty intersection with $D_x \doteq \{d \in D : \varphi(v) \leq x\}$.*

By processing in this manner the above family of measuring functions we get a set of "size function tables" whose mean values collected together finally give the feature vector.
We have chosen a 6-line sheaf to compute the measuring functions. As a test for our feature combination method we have considered each single component of the vector as a *distinct* feature and built a separate FOD for it. This allowed us to choose a low number of states for each frame and show how our system can reject incorrect measurements.

6.2. Performances

The PantoMouse planar workspace is a 19×19 mm square: we have first chosen a sample trajectory composed by a regular mesh of points with $\epsilon = 1$ mm. The corresponding feature vector evolution has shown that measurements were smooth enough to be discerned by this approximation and suggested the following assignment for the HMM's number of states: $n_1 = 2,\ n_2 = 2,\ n_3 = 2,\ n_4 = 3,\ n_5 = 3,\ n_6 = 2$. After building the family of discrete feature spaces and the refining maps to $\tilde{\mathcal{Q}}$ we made the robot execute a number of movements for testing our tracking algorithm. Figure 2a compares one of these motions to the estimated positions obtained by encoding feature data as Bayesian belief functions, combining them and then computing the relative plausibility of each sample position as we have seen in Section 4d. The resulting estimation error was 2.3585.
It is interesting to see how a increased cardinality of the approximate feature spaces can improve the estimation precision. Figure 2b shows the effect of refining the frames related to the second and third component of the feature vector from 2 to 3 elements. The estimation error $\bar{e} = 1.9244$ confirms our visual impression. It should wonder how these low-state feature spaces can produce such a nice behaviour of the system.

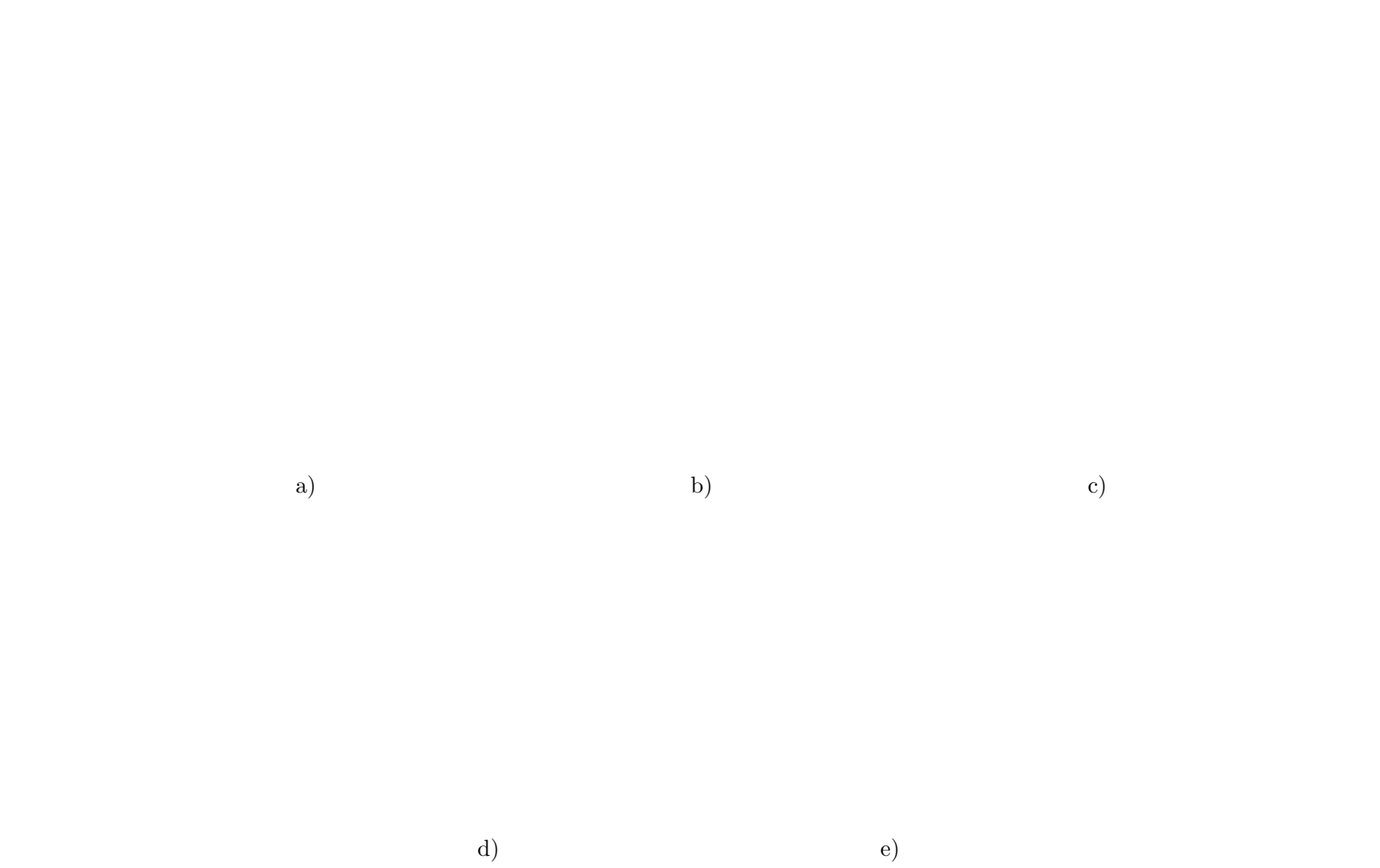

Figure 2. a) Real (gray) and estimated (black) positions of the robot in a tracking example. b) Tracking results with $n_2 = n_3 = 3$. c) Effect of a sparse sample trajectory. d) Effect of consonant measurement functions. e) Precision increment due to features addition.

We have built also a second evidential model for PantoMouse by choosing a more sparse trajectory obtained by sampling the first one. The estimation error has grown to $\bar{e} = 2.6579$ as Figure 2c shows for the new $\tilde{\mathcal{Q}}$ could not cope with high feature variation rates.
The choice of the b.f. constructor can affect the behaviour of the estimation process too. The estimates resulting from the combination of consonant measurement functions are shown in Figure 2d. It can be observed that they are not significantly different from what we obtain by using Bayesian b.f.'s ($\bar{e} = 2.0348$).
Finally we introduced another kind of feature to show how the addition of new information improves the estimate. In particular we calculated the baricentre of the images thought as intensity functions and updated the evidential model adding two other scalar frames with $n_7 = 3$ and $n_8 = 2$ respectively.
Figure 2e illustrates how the updated model better describes the PantoMouse motion: the estimation error becomes $\bar{e} = 1.6788$. Nevertheless two estimate clusters around the points (4.3,3) and (6.5,14) tell us that the feature set we have chosen is not informative enough to distiguish positions inside these regions. Building a more refined FOD for the horizontal component of the baricentre could be useful.
During all these tests the conflict graph has always resulted completely connected: each pair of belief functions has shown a very low level of conflict. This is not surprising for we have chosen low values for the n_i's so the measurement functions can hardly support contradictory propositions. Adapting this method to more significant applications will require studying the interesting *tradeoff* between estimation precision and measurement conflict.

7. CONCLUSIONS AND FUTURE WORK

In this paper we have analyzed a way of integrating different sources of information in order to produce a robust estimate of the configuration of an articulated body. We adopted the evidential reasoning as the most natural theoretical framework in which a solution to this problem can be found. We have shown how to combine the hidden

Markov models implicit quantization mechanism and the idea of refinings between compatible frames for building an evidential model of the object. In this environment the nature of the relationships among completely different representations finds a formal definition.
The consistency of distinct viewpoints of object motion is described by the notion of "conflict level": a learning algorithm allowing us to detect the most coherent group of features has been formulated. The influence of the number of states of the discretized feature spaces on tracking precision has been analyzed: we have shown how relatively rough discretizations can give appreciable results.
Several theoretical properties of the system should be investigated and the relationships among configuration space, sample trajectory, continuous and discrete feature spaces will be analyzed in the future. In particular, the pointwise estimation mechanism must be checked in more complex situations like high-dimensional non-linearly connected parameter spaces. Interesting possibilities involve the integration of our feature-based framework with analytical model-based approaches[12][13] to object tracking. The knowledge of the kinematic model of the object, for example, could lower the estimate sensitivity to the detail level of feature representations.

8. ACKNOWLEDGMENTS

We wish to thank the Industrial Engineering Laboratory of our department for the use of their PantoMouse robot in our experiments. We would like to cite in particular Roberto Oboe and Francesca Bettini for their gentle collaboration.

REFERENCES

1. Aaron F. Bobick and Andrew D. Wilson, *Learning visual behavior for gesture analysis*, IEEE Symposium on Computer Vision, November 1995.
2. Fabio Cuzzolin and Ruggero Frezza, *Using hidden Markov models and dynamic size functions for gesture recognition*, Proc. of the 8th British Machine Vision Conference (BMVC97), September 1997.
3. A.P. Dempster, *Upper and lower probabilities induced by a multivariate mapping*, Annals of Mathematical Statistics **38** (1967), 325–339.
4. R. Elliot, L. Aggoun, and J. Moore, *Hidden Markov models: estimation and control*, 1995.
5. P. Frosini, *Measuring shape by size functions*, Proc. of SPIE on Intelligent Robotic Systems, vol. 1607, 1991, pp. 122–133.
6. Michel Grabisch, Hung T. Nguyen, and Elbert A. Walker, *Fundamentals of uncertainty calculi with applications to fuzzy inference*, Kluwer Academic Publishers, 1995.
7. H.T. Hestir, H.T. Nguyen, and G.S. Rogers, *A random set formalism for evidential reasoning*, Conditional Logic in Expert Systems, North Holland, 1991, pp. 309–344.
8. J. Kohlas and P.A. Monney, *Theory of evidence - a survey of its mathematical foundations, applications and computational anaylsis*, ZOR- Mathematical Methods of Operations Research **39** (1994), 35–68.
9. Jurg Kohlas and Paul-Andr Monney, *A mathematical theory of hints - an approach to the dempster-shafer theory of evidence*, Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, 1995.
10. H.T. Nguyen, *On random sets and belief functions*, J. Mathematical Analysis and Applications **65** (1978), 531–542.
11. D. Pagac, E.M. Nebot, and H. Durrant-Whyte, *An evidential approach to map-building for autonomous vehicles*, IEEE Trans. on Robotics and Automation **14, No 4** (1998), 623–629.
12. James M. Rehg and Takeo Kanade, *Digiteyes: Vision-based human hand tracking*, Tech. report, School of Computer Science, Carnegie Mellon University, CMU-CS-93-220, December 1993.
13. _____, *Visual tracking of self-occluding articulated objects*, Tech. report, School of Computer Science, Carnegie Mellon University, CMU-CS-94-224, December 1994.
14. Glenn Shafer, *A mathematical theory of evidence*, Princeton University Press, 1976.
15. T. M. Strat, *Continuous belief functions for evidential reasoning*, Proceedings of the National Conference on Artificial Intelligence (Institute of Electrical and Electronical Engineers, eds.), August 1984, pp. 308–313.