Framework for Multisensory Convergence

Miroslav Kárný[*], **Kevin Warwick**[†], **Tatiana V. Guy**[*], **Jan Kracík**[*]

[*]Adaptive Systems Department
Institute Information Theory and Automation
P.O. Box 18, 182 08 Prague, Czech Republic
e-mail: school@utia.cas.cz

[†] Department of Cybernetics
University of Reading
Whiteknights, Reading, RG6 6AY, UK

Abstract. In this paper we consider the multisensory convergence problem, that is when signals from several different sensors report on the same event, possible through the employment of different pieces and types of information. A unified approach based on non-linear filtering is proposed here. Firstly, Kalman filtering is applied to a simple linear state-space model. The filter takes input data from different sensors and produces estimate of the physical state describing, for example, environment of a robot. This shows structure of optimal sensor fusion. Then, a general version of such a filtering is presented with a careful account for computational complexity. This leads to modelling through finite probabilistic mixtures: both sensors and global state estimates are modelled by finite mixtures of uni-modal probability density functions (pdf). Bayesian non-linear estimation is then used to mix together the different sensor responses to a single information driven estimate of environment state. The proposed theory will be confronted with conditions offered by a prototype 5-sensory robot head that has been set up as a test-bed platform for the multisensory fusion algorithms.

1 Introduction

In the biological world the ability of a creature to integrate information obtained from a number of sensory mechanisms at its disposal in order to help it decide what action to take at any instant in time is paramount in its whole being. Because of such input a creature is able to orient itself in space, in relation to its perception of objects within a region of that space, respond to changes perceived in the environment, for such as nutrition, and perhaps, important of all, preserve itself as a dependable life form. In any search for artificially alive beings, autonomous devices that exist for long periods under their own adaptive control or even simply robots that interact with humans, efficient and functional multisensory fusion is an area of primary concern. Yet it is an area of research which has received relatively little attention to date, other than for a specific problem solution for a robot faced with two pieces of information and a quick decision to make e.g. (Engels & Schoner 1995, Tyrrell 1994, Foresti & Regazzoni 2002, Jassemi-Zourgani & Necsulescu 2002).

In particular, in any one biological system, operative sensors can be extremely different in their characteristics, they can operate for example with different time constants and even focus on different aspects of the environment. As has been pointed out (Howard 1998) sensors can operate with respect to each other, in a series, nested or parallel fashion, in particular with the possibility of being arranged in a "multi-cue system" in order to provide information regarding the same event. Such information may either be redundant (competitive) or complementary. Combining such complementary information can then be employed as a necessary part of an overall control mechanism, as indeed appears to be the case with retina stabilization in humans (Robinson 1977) in which high frequency components of head rotation, sensed by the vestibular system, are mixed with low frequency optokinetic signals. Some attempts at multisensory integration (e.g. (Boss, et al. 2001)) have fused multisensory signals, in a theoretical environment by quite simply taking a weighted average of the two inputs involved. This has the advantage of dealing with redundancy when one input proves to be problematic or even becomes dysfunctional however it is far removed from biological reality (van Beers, et al. 99) and generally such occurrences can be dealt with by a straightforward hierarchical decision mechanism as that in the human vision system (Wolfe 1994). Indeed it is acknowledge in (Boss et al. 2001) that a straightforward averaging mechanism can only work if it is, as they put it, supported by a "book keeping" stage, which has the ability to reconfigure the averaging process accordingly! Clearly what is needed is a more firmly based theoretical approach of sensor fusion that exhibits biological plausibility whilst proving to be practically feasible and operable in a robotic environment. It is, we believe, that which we are suggesting and striving for here.

Evidence suggests (Stein & Meredith 1993) that neurons in the superior colliculus of the brain form a sort of sensory map of the surroundings based on a combination of auditory and visual information. An overlap appears to exist between the neurons involved, i.e. a high proportion of them are truly multisensory. As a result it is often the case, when stimuli of different modality are complementary, that the overall neural response is larger than the sum of the uni-modal responses. Conversely where competitive instances occur overall neural activity can even decrease. One interesting feature is that the response enhancement is strongest when complementary signals are weak and is less apparent as such signals become uni-modally stronger.

It is with regard to the biological observations discussed that we wish to aim for a more plausible and practically operable overall sensor fusion approach. Whilst taking inspiration from the biological world to some extent, it is not felt to be a sensible approach, at this time, simply to throw further biologically inspired techniques at the problem and hope that it all "comes out in the wash", such would be the case with neural networks. What we need to gain, given the problem in hand, is a deeper understanding of what possibilities there are and to technically assess what is possible.

1.1 Framework, Layout and Presentation Way

In this paper we consider non-linear-filtering framework (Jazwinski 1970, Anderson & Moore 1979) with which to tackle the multisensory convergence problem. Applicability of this conceptually plausible approach depends strongly on availability of such observation and time-evolution models that can efficiently processed. Finite mixtures of uni-modal probability density functions (pdfs) (Titterington, et al. 1985, McLachlan & Peel 2001) seem to be adequate. It seems to be both biologically plausible (Berthouse & Kunniyoshi 1998) and has been successfully applied in audiovisual object tracking (Beal & Jojic 2003). Still the presented results depend too much on the specific problem and give a weak guide how to solve other types of sensor convergence.

The advocated approach is a relatively straightforward elaboration of stochastic filtering theory. Two seemingly different ways were originally inspected (i) joint state modelling of sensors and (ii) global modelling of the observed environment. At the end we found that the second option results in a specific version of the first one. It is demonstrated on intentionally simplified case in section 2. It shows that the state-space formulation has the added advantage of mathematically taking into account the correlations that exist between different sensors, directly in the formation of the state vector. Subsequently it is the separate state vector elements that we wish to fuse, rather than highly correlated, noise ridden, inaccurate measured sensor values.

The general discussion is in section 3. It is shown that computational obstacles caused by complexity of general stochastic filtering – that performs formally optimal sensor fusion – can be to significant extent suppressed. The key step is modelling of local sensors by probabilistic mixtures with local components and a flat background component. It leads naturally to the use of finite mixtures (Titterington et al. 1985, Quinn, et al. 2003) for description of global estimate, too. Feasible treatment of both data updating and time updating of estimates of high-dimensional field of physical states of the environment state are proposed.

The proposed theory will be tested on 5 sensory robot head that is described in section 4. The head can be goal directed, based on its multisensory input and can be given simple task, for instance, the head is to remain a fixed distance away from a (potentially moving) object which it can track through the variety of sensors at its disposal. The foreseen tests with it are briefly commented in section 5

1.2 Notation

In the exposition, the following notation is used: $\equiv$ definition by equality; a^* set of values of variable a; $a^{\mathcal{C}}$ complement of the set a^*, $f(a|b)$ probability density function (pdf) of a conditioned on b; $\propto$ proportionality; $t \in t^* \equiv \{1, 2, \ldots, \mathring{t}\}$ discrete time, always placed as the last subscript; q position in (possibly phase) space q^*; $s \in s^*$ sensor indices; $q^*_{s;t}$ observation range of the sth sensor at time t; $x_{q;t}$ physical state of the observed environment at position q and time t (a finite dimensional vector); $x_t(q^*)$ collection of values $x_{q;t},\ q \in q^*$, at time t; $x(t, q^*)$ collection of values $x_{q;\tau},\ q \in q^*$, at time moments $\tau \leq t$; $y_{s;t}$ (vector) output of the sensor $s \in s^*$, at time t; $u_{s;t}$ (vector) exogenous input to the sensor $s \in s^*$, at time t.

2 State-Space Description of Sensing

This section introduces gently into the problem and outlines the structure of the solution. Multiple sensors operating in high dimensions and their application within a robotics environment are discussed subsequently.

We consider two separate and potentially different sensors, labelled by $s \in s^* \equiv \{1, \mathring{s} \equiv 2\}$ reporting

on the same scalar physical quantity x_t at discrete time $t \in t^* \equiv \{1, \ldots, \mathring{t}\}$ at the same spatial position $q \in q^*$. Outputs of sensors $y(\mathring{s}, t) \equiv [y_{1;1}, \ldots, y_{1;t}, y_{2;1}, \ldots, y_{2;t}]$ observed till time t should be converted into estimate $\hat{x}_{t|t}$ of the physical state x_t. Design of the estimator processing outputs of several sensors is referred to as the multisensory convergence problem (Meredith 2002) or sensor fusion.

Models of sensors relate their outputs $y_t(\mathring{s})$ to the physical state x_t. Here, linear static models are assumed:

$$y_{s;t} = h_s x_t + e_{ys;t},\ s \in s^* \Leftrightarrow y_t(\mathring{s}) = [h_1, h_2]x_t + e_{y;t}(\mathring{s}), \tag{1}$$

where the known coefficient h_s characterize sensors and unobserved $e_{y;t}$ is white measurement noise.

Time evolution of the physical quantity is modelled by linear model:

$$x_t = ax_{t-1} + e_{x;t} \tag{2}$$

with a known coefficient a (degenerate version of state matrix) and white process noise $e_{y;t}$.

The estimate $\hat{x}_{t|t}$ is gained through a filter that fusses all available measurements into the posterior pdf $f(x_t|\mathcal{P}_t),\ \mathcal{P}_t \equiv y(\mathring{s}, t)$. This pdf describes belief into possible values of the estimated x_t. Its mean or mode serves as the point estimate $\hat{x}_{t|t}$ usually searched for and variance determines confidence of this estimate.

Under appropriate conditions, the conditional pdf has a fixed functional form determined by a finite-dimensional information state V_t that evolves according to the state equation of the fussing filter, (Jazwinski 1970):

$$V_{t|t} = \mathcal{V}_1(V_{t|t-1}, y_t(\mathring{s})),\ \ V_{t+1|t} = \mathcal{V}_2(V_{t|t}) \tag{3}$$

and generates the desired estimate through output equation of the filter:

$$\hat{x}_{t|t} = \mathcal{X}(V_{t|t}). \tag{4}$$

The functions determining the state evolution $\mathcal{V}(\cdot) \equiv [\mathcal{V}_1(\cdot), \mathcal{V}_2(\cdot)]$ and output of the filter $\mathcal{X}(\cdot)$ are determined by models of sensors (1), by the model of the physical state evolution (2) and by the definition of the state estimate required. Considering whiteness and normality of the involved noises, the conditional pdfs preserves normal form. Its moments evolves according to the celebrated Kalman filter, (Jazwinski 1970):

Data updating

$$\begin{aligned} \mathcal{V}_1(\cdot):\quad & \hat{x}_{t|t} = \hat{x}_{t|t-1} + \left(y_t(\mathring{s}) - h\hat{x}_{t|t-1}\right) \underbrace{h' \left(h'hP_{t|t-1} + \mathrm{cov}[e_{y;t}(\mathring{s})]\right)^{-1}}_{\text{Kalman gain} \equiv g_{t|t}} \\ & P_{t|t} = (1 - g_{t|t}h')P_{t|t-1} \end{aligned} \tag{5}$$

Time updating

$$\begin{aligned} \mathcal{V}_2(\cdot):\quad & \hat{x}_{t+1|t} = a\hat{x}_{t|t} \\ & P_{t+1|t} = a^2 P_{t|t} + \mathrm{var}(e_{x;t}). \end{aligned}$$

The function $\mathcal{X}$ just selects $\hat{x}_{t|t}$ as the required estimate. The sufficient statistics are the conditional mean $\hat{x}_{t|\tau}$ and variance $P_{t|\tau}$ of $f(x_t|\mathcal{P}_\tau),\ \mathcal{P}_\tau \equiv y(\mathring{s}, \tau)),\ \tau \in \{t-1, t\}$ and $'$ is transposition. The two-dimensional Kalman gain $g_{t|t}$ is deterministic function of noise characteristics and takes into account possible correlations between involved noises. The second term in the first equation of (5) shows how the sensors outputs are fused in this case. It is worth stressing that the evolved information state $V_{t|\tau} \equiv [\hat{x}_{t|\tau}, P_{t|\tau}]$ describes uni-modal pdf that can be interpreted as activity distribution of the fused sensors. Such a uni-modal pdf has biological interpretation as a target object being mapped by means of a retinotopical projection (Boss et al. 2001).

In our case however the pdf results from operational characteristics of sensors and considered evolution of the observed physical state. A similar view point can be found in biologically inspired neural representations of sensing stimuli. They exhibit extended receptive fields, effectively blurring the physical state on the sensor input. In essence it is assumed that an activity at a location q in a sensory input drives the state vector through a bell shaped distribution.

The above filtering structure can be also described in terms of observers (Warwick 1987) and extended to non-linear sensor and time-evolution models and multi-dimensional case (Jazwinski 1970). Computational feasibility of the resulting filter is decisive for applicability of any such extension. The subsequent section provide a relatively general solution while preserving the simple reasoning structure outlined above.

3 Probabilistic Mixing Sensory Data

References and results of section 2 indicate the expected structure of fusing sensory data. This section tries to embed the problem into a probabilistic framework that allows us to see the problem structure, to specify more precisely the involved elements and operations and propose a general solution.

At each $q \in q^* \equiv$ *position (possibly phase) space* and at each time instant t, the inspected environment is characterized by a finite-dimensional state vector $x_{q;t}$ of *physical quantities*. They are observed by several sensors. Their data serve for estimation of the state field $x_t(q^*)$ over whole space q^*. The estimation is formulated and solved as stochastic filtering recalled in section 3.1. Sensor models are discussed in subsection 3.2. Data updating of the physical-state-vector estimates by single sensor is discussed in subsection 3.3. Joint data updating that fuses data of several sensors into a common estimate of physical state is presented in subsection 3.4. Then, models of sensors belonging to tractable dynamic exponential family (DEF) (Barndorff-Nielsen 1978) are discussed with their use for data updating of conjugate prior pdf, subsection 3.5. Dimensionality problem present even in this nice family of models is addressed through mixture-based re-parameterization presented in subsection 3.6. Global estimate of the physical states $x(q^*)$ over whole considered environment is discussed in subsection 3.7 where its data updating is also discussed. Time evolution of these estimates is resolved in subsection 3.8 under simplifying assumption of slow temporal changes of physical state of the environment. This assumption is justified by a relatively high rate with which the sensors can inform us on the current state of the environment. The section is concluded by summarizing the overall fusion (filtering) algorithm, subsection 3.9. Questionable steps are discussed here, too.

3.1 Stochastic filtering

Let $X_t \equiv x_t(q^*)$ be a state to be estimated using measurement data $D_t \equiv d_t(\mathring{s}) \equiv [d_{1;t}, \ldots, d_{\mathring{s};t}]$, where data $d_{s;t}$ related to s-th sensor consist of the measured sensor output $y_{s;t}$ and possibly of external sensor inputs $u_{s;t}$.

Let us assume that *observation model* $f(D_t|\mathcal{P}_{t-1}, X_t)$ and *time evolution model* $f(X_{t+1}|\mathcal{P}_t, X_t)$ are at disposal. The symbols $\mathcal{P}_{t-1},\ \mathcal{P}_t$ used in these conditional pdfs, denote the information processed up to and including time moments $t-1,\ t$, respectively.

Stochastic filtering updates posterior pdf $f(X_t|\mathcal{P}_t)$ of the unknown state X_t. It is described by the coupled formulas, (Jazwinski 1970, Peterka 1981) (Kalman filter discussed in section 2 is its special version):

$$\text{Data updating} \quad f(X_t|\mathcal{P}_t) \propto f(D_t|\mathcal{P}_{t-1}, X_t) f(X_t|\mathcal{P}_{t-1}) \tag{6}$$

$$\text{Time updating} \quad f(X_{t+1}|\mathcal{P}_t) = \int f(X_{t+1}|\mathcal{P}_t, X_t) f(X_t|\mathcal{P}_t)\, dX_t. \tag{7}$$

The recursion starts with the externally supplied prior pdf $f(X_1|\mathcal{P}_0)$.

Remarks

1. Filtering processes optimally outputs of all sensors, i.e. fusses them optimally. Optimality means that all available information about the estimated state is exploited.
2. Specific models and approximations used in the functional recursion (6), (7) decide on the applicability.
3. Any reasonable characteristic of $f(X_t|\mathcal{P}_t)$, e.g. mean or mode, can be selected as a point estimate of X_t.
4. A pair of time indices occur in (6), (7). One refers to the estimated variable X_t, the other one to the information $\mathcal{P}_{t-1}$ or $\mathcal{P}_t$ included into the condition. In order to stress it, we shall use the subscript $t|t-1$ or $t|t$ at time positions whenever needed.
5. Recursive form of filtering implies that the posterior pdf $f(X_t|\mathcal{P}_{t-1})$ serves as a prior one for the time t.

3.2 Models of Sensors

Sensors are man-made imprecise devices of a known structure. Output $y_{s;t}$ of the sensor $s \in s^* \equiv \{1, \ldots, \mathring{s}\}$ at time t is corrupted by sensor dynamics, its non-linearity and measurement noise. General description of the sensor at a specific position is given by the conditional pdf $f(y_{s;t}|y_s(t-1), x(t, q^*))$.

Sensing is always to some extent *local*. It reflects some entries of physical state x of the environment distributed only on a subset $q^*_{s;t} \subset q^*$ of the space q^*. The subset varies with variations of sensor positions.

Sensing dynamics can be modelled by a dependence of the current sensor output $y_{s;t}$ on *regression vector* $\psi_{s;t}$ consisting of several delayed sensor outputs $y_{s;\tau}$, $\tau = t-1, t-2, \ldots, t-\partial_y$ and (possibly) on current and delayed *exogenous sensor input* $u_{s;\tau}$, $\tau = t, t-1, \ldots, t-\partial_u$. The exogenous sensor input is supposed to meet

natural conditions of control, NCC, (Peterka 1981), i.e. it uses at most the information about the environment state contained in the measured data. Thus, the sensor approximately exhibits the Markov property:

$$f(y_{s;t}|u_s(t), y_s(t-1), x(t,q^*)) = f(y_{s;t}|\psi_{s;t}, x_t(q^*_{s;t})). \tag{8}$$

The pdf (8) is either implied directly by the construction of the sensor or can be gained through the Bayesian estimation of its parameters (Peterka 1981).

3.3 Exploitation of Sensor Reading

Given a prior pdf on the environment state $f(x_t(q^*)|\mathcal{P}_{t-1})$, the measurement of s-th sensor refreshes it to:

$$f(x_t(q^*)|\mathcal{P}_{s;t}) \propto f(y_{s;t}|\psi_{s;t}, x_t(q^*_{s;t}))f(x_t(q^*)|\mathcal{P}_{t-1}) \tag{9}$$

where $\mathcal{P}_{s;t} \equiv (d_{s;t}, \mathcal{P}_{t-1}) \equiv (y_{s;t}, u_{s;t}, \mathcal{P}_{t-1})$ is the information $\mathcal{P}_{t-1}$, acquired up to time $t-1$ from all sensors, updated by s-th sensor data only. The relation (9), performing a version of data updating (6), is just a plain Bayes rule valid under NCC.

Decomposing $q^* = q^*_{s;t} \cup q^{\mathcal{C}}_{s;t}$, with the *complement of* $q^*_{s;t}$ defined $q^{\mathcal{C}}_{s;t} \equiv q^* \setminus q^*_{s;t}$, and using chain rule, we can re-write (9) into the form:

$$\begin{aligned} &f\left(x_t\left(q^{\mathcal{C}}_{s;t}\right)|x_t(q^*_{s;t}), \mathcal{P}_{s;t}\right) f(x_t(q^*_{s;t})|\mathcal{P}_{s;t}) \propto \\ &\propto f(y_{s;t}|\psi_{s;t}, x_t(q^*_{s;t}))f\left(x_t\left(q^{\mathcal{C}}_{s;t}\right)|x_t(q^*_{s;t}), \mathcal{P}_{t-1}\right) f(x_t(q^*_{s;t})|\mathcal{P}_{t-1}). \end{aligned} \tag{10}$$

Integrating (10) over $x_t\left(q^{\mathcal{C}}_{s;t}\right)$, we get:

$$f(x_t(q^*_{s;t})|\mathcal{P}_{s;t}) \propto f(y_{s;t}|\psi_{s;t}, x_t(q^*_{s;t}))f(x_t(q^*_{s;t})|\mathcal{P}_{t-1}), \tag{11}$$

i.e. the information about state in the neighborhood $q^*_{s;t}$ is updated by the Bayes rule irrespectively of the information on the state on its complement $q^{\mathcal{C}}_{s;t}$. By inserting this result into (10), we see that:

$$f\left(x_t\left(q^{\mathcal{C}}_{s;t}\right)|x_t(q^*_{s;t}), \mathcal{P}_{s;t}\right) = f\left(x_t\left(q^{\mathcal{C}}_{s;t}\right)|x_t(q^*_{s;t}), \mathcal{P}_{t-1}\right), \tag{12}$$

i.e. this conditional pdf of x_t on the complement $q^{\mathcal{C}}_{s;t}$ of the set $q^*_{s;t}$ is unchanged by the sensor output $y_{s;t}$.

This intuitively appealing result is practically important as the information change caused by the local sensor measurement causes a local change of $f(x(q^*)|\mathcal{P}_{t-1})$, which is extremely high-dimensional pdf defined over the environment states in all positions $q \in q^*$.

3.4 Fusion of Sensor Readings

Probabilistic fusion of several sensor reading is conceptually straightforward if their common model $f(y_t(\mathring{s})|y(\mathring{s}, t-1), u(\mathring{s}, t), x(t, q^*))$ is available. Sensors are, however, independently constructed devices whose outputs are correlated just due the overlapping observations of the environment physical state, see Figure 1. Thus, *without loss of generality*, their outputs can be assumed to be *conditionally independent*, i.e.:

$$\begin{aligned} f(y_t(\mathring{s})|y(\mathring{s}, t-1), u(\mathring{s}; t), x(t, q^*)) &= \prod_{s \in s^*} f(y_{s;t}|y_s(t-1), u_s(t), x(t, q^*)) \underbrace{\equiv}_{(8)} \\ &\equiv \prod_{s \in s^*} f(y_{s;t}|\psi_{s;t}, x_t(q^*_{s;t})). \end{aligned} \tag{13}$$

With this joint description of sensors, the fusion of their readings reduces simply to a straightforward application of the Bayes rule:

$$\begin{aligned} f(x_t(q^*)|\mathcal{P}_t) &\propto f(y_t(\mathring{s})|y(\mathring{s}, t-1), u(\mathring{s}, t), x_t(q^*))f(x_t(q^*)|\mathcal{P}_{t-1}) = \\ &= \prod_{s \in s^*} f(y_{s;t}|\psi_{s;t}, x_t(q^*_{s;t}))f(x_t(q^*)|\mathcal{P}_{t-1}). \end{aligned}$$

Arguments similar to that for (12) imply that the high-dimensional prior pdf $f(x_t(q^*)|\mathcal{P}_{t-1})$ is modified on:

$$x\left(q^*_t(\mathring{s})\right) \equiv x\left(\cup_{s \in s^*} q^*_{s;t}\right) \tag{14}$$

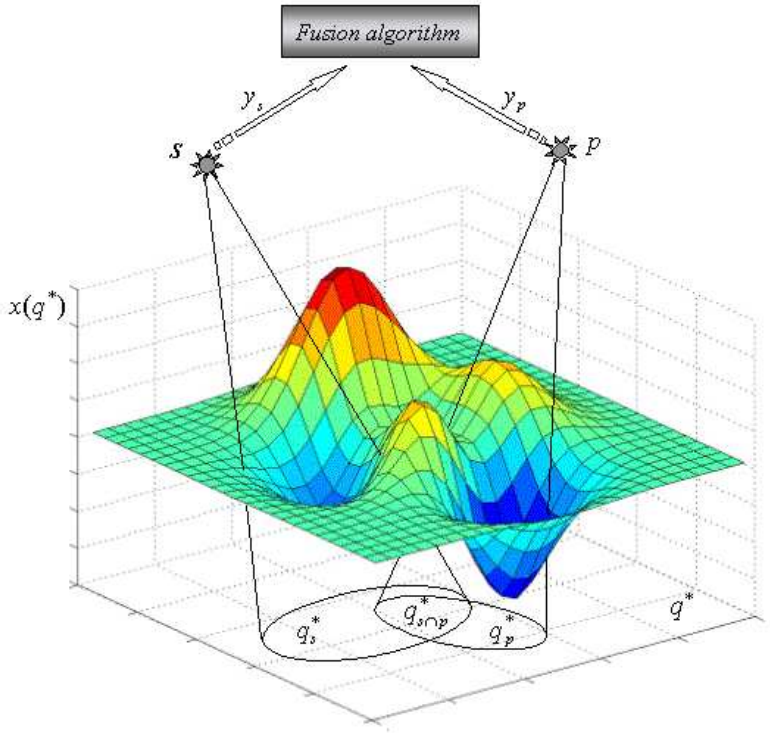

Figure 1. Multi-sensors with overlapping observed positions.

Remarks

1. The marginalization of the global pdf $f(x_t(q^*)|\mathcal{P}_{t-1})$ is the most time consuming operation. It need not be done unless sensor positions are changed. The remaining operations are computationally cheap as they work on – always very localized – position sets q^*_s defined by the used sensors.
2. The outlined data updating maps $f(x_t(q^*)|\mathcal{P}_{t-1})$ on $f(x_t(q^*)|\mathcal{P}_t)$. It is necessary to specify the time-updating mapping

   $$f(x_t(q^*)|\mathcal{P}_t) \to f(x_{t+1}(q^*)|\mathcal{P}_t)$$

   in order to complete the learning recursion. Generally, it needs the model of time evaluation $f(x_{t+1}(q^*)|x_t(q^*), \mathcal{P}_t)$, see (7). It is hard to get it even in much simpler situations and the non-linear stochastic filtering, subsection 3.1, is solvable in rare cases (Daum 1988). Thus, at the considered level of generality, we have to restrict ourselves to them or to the case slowly varying values of $x_t(q^*)$ and apply a version of forgetting (Kulhavý & Zarrop 1993). The latter version is adopted and discussed below.

3.5 Sensor Models in Dynamic Exponential Family

Inherent problem dimensionality forces us to approximate sensor models by pdfs in *dynamic exponential family (DEF)*. The parameterized model of a sensor (8) belongs to DEF iff it is described by the formula:

$$f(y_{s;t}|\psi_{s;t}, x_t(q^*_{s;t})) = \exp\left\langle B_s(\Psi_{s;t}), H_s C(x_t(q^*_{s;t}))\right\rangle, \; s \in s^*, \text{ where} \tag{15}$$

data vector $\Psi_{s;t} \equiv [y_{s;t}, \psi_{s;t}]$ = [sensor output,regression vector];
the data vector $\Psi_{s;t}$ can be updated recursively, i.e. there is a function $\tilde{\Psi}_s$ such that $\Psi_{s;t} = \tilde{\Psi}_s(d_{s;t}, \Psi_{s;t-1})$ with the *data item* $d_{s;t} \equiv [y_{s;t}, u_{s;t}]$ =[sensor output, exogenous sensor input];
H_s is a fixed *linear selector* of x_q entries that are reflected in outputs of s-th sensor;
$\left\langle B_s(\Psi_{s;t}), H_s C(x_t(q^*_{s;t}))\right\rangle$ is a scalar product (a bilinear functional) of arrays $B_s(\cdot)$, $H_s C(\cdot)$, for which there is a *conjugated linear selector* $\bar{H}_s$ such that

$$\left\langle B_s(\Psi_{s;t})\bar{H}_s, C(x_t(q^*_{s;t})\right\rangle = \left\langle B_s(\Psi_{s;t}), H_s C(x(q^*_{s;t}))\right\rangle,$$

finite-dimensional functions $B_s(\cdot), H_sC(\cdot)$ of respective arguments are arranged into arrays of compatible dimensions,

the array $C(x(q^*_{s;t}))$ has always a constant (unit) entry: this allows us to include normalizing factor of the pdf as an entry of B.

Existence of *self-reproducing prior pdf*, conjugated with the model of the sensor in DEF, determined by a finite-dimensional sufficient statistics, explains its practical significance. DEF allows, moreover, to update the statistics recursively.

Under the assumption (15), the conjugated prior pdf has the form, see. (14):

$$f(x_t(q^*_t(\mathring{s}))|\mathcal{P}_{t-1}) \propto \prod_{s\in s^*} \exp\left\langle V_{s;t|t-1}, H_sC(x_t(q^*_{s;t}))\right\rangle f\left(x\left(q^{\mathcal{C}}_t(\mathring{s})\right)|x\left(q^*_t(\mathring{s})\right), \mathcal{P}_{t-1}\right), \tag{16}$$

where $V_{s;t|t-1}$ are individual sufficient statistics defining this prior pdf. The non-updated pdf on the complement $q^{\mathcal{C}}_t(\mathring{s}) \equiv q^* \setminus q^*_t(\mathring{s})$ can be formally arbitrary. Practically, it has to be carefully chosen to allow practical implementation of the full learning cycle consisting of data and time updating of the global estimate $f(x_t(q^*)|\mathcal{P}_{t-1})$. In any case, the updating by data from given sensors reduces to algebraic recursions:

$$V_{s;t|t} = V_{s;t|t-1} + B_s(\Psi_{s;t})\bar{H}_s,\ s \in s^* \tag{17}$$

and the posterior pdf $f(x_t(q^*)|\mathcal{P}_t)$ has the form (16) with $V_{s;t|t}$ replacing $V_{s;t|t-1}$.

3.6 Re-Parameterization of Sensor Models

The presented straightforward extension of the known data updating in DEF has significant drawback in the sensor convergence problem: the union of the spaces $q^*_t(\mathring{s})$ observed by sensors evolves with time. A correct determination of the conjugated prior pdf on this set requires computer intensive manipulations with the joint pdf $f(x_{t+1}(q^*)|\mathcal{P}_t)$. Its even approximate description has to be very detailed and consequently extremely high-dimensional. Otherwise, it is impossible to store fine "traces" of sensing made in past. The re-parameterization of the sensor models, proposed here, avoids this problem to a substantial degree. At the same time, the proposed solution seems to be both general and flexible enough.

Essentially, we re-parameterize sensor models so that they work formally on whole time-invariant space q^*. An individual sensor is modelled by a mixture of the sharp component $f_s(y_{s;t}|\psi_{s;t}, x_t(q^*))$ in DEF, that has practical support on $x(q^*_{s;t})$, and of a flat $\mathcal{U}(x(q^*))$ component (uniform when q^* is finite):

$$\begin{aligned} f(y_{s;t}|\psi_{s;t}, x(q^*)) &= \alpha_s f_s(y_{s;t}|\psi_{s;t}, x_t(q^*)) + (1-\alpha_s)\mathcal{U}(x(q^*)) \equiv \\ &\equiv \alpha_s \prod_{q\in q^*} \exp\left\langle B_s(\Psi_{s;t}), H_sC(x_{q;t})\right\rangle + (1-\alpha_s)\mathcal{U}(x(q^*)). \end{aligned} \tag{18}$$

The function $C(x_{q;t})$ of the local state $x_{q;t}$ is chosen as sensor independent. The sensor specificity is assumed to be concentrated into the selector H_s. The probabilistic weight $\alpha_s \in (0,1)$ is known (estimated) characteristics of the sensor chosen so that on the set $q^*_{s;t}$ the contribution of $\alpha_s f_s(y_{s;t}|\psi_{s;t}, x_t(q^*))$ is higher than that of $(1-\alpha_s)\mathcal{U}(x(q^*))$ and vice versa on the complement $q^{\mathcal{C}}_{s;t}$.

Remarks

1. The mixture models of sensors (18) are considered further on. Of course, specific physical knowledge may lead to other, often simpler, types of models.
2. Construction of the sensors often guarantees that the locality of sensing, reflected in data updating (11), (12), is approximately met when using the model (18) in Bayes rule.
3. Physical states $x_{q;t}$ considered on the whole space q^* are assumed to enter the model through $\sum_{q\in q^*} C(x_{q;t})$. This assumption can be generalized to $\sum_{q\in q^*} \beta_q C(x_{q;t})$, with fixed weights β_q. These weights are supposed to reflect dependence of physical properties on sensor position q. The states $x_{q;t}$ enter the model with weights respecting physical properties of the observed environment. Presentation simplicity makes us to consider the version without weighting.

3.7 Global Estimate and its Data Updating

The overall state space $x(q^*)$ is extremely large and cannot be efficiently described in an exact point-wise manner. At the same time, the dimensionality implies that the number of areas with interesting configurations

of physical states is very limited. Moreover, these areas represent as a rule "objects" with a physical meaning and occupy connected parts of the geometrical space. These parts can be further split into a finite amount of subsets on which physical states have very similar values at the inspected time moment.

Let us suppose, that each region with similar physical states is described by so called *component*. It implies that the joint prior pdf can be approximated by a finite mixture with (relatively sharp) components formed by pdfs conjugated to DEF and by a flat (uniform) pdf:

$$f(x_t(q^*)|\mathcal{P}_{t-1}) = \sum_{c=1}^{\mathring{c}_{t|t-1}} \alpha_{c;t|t-1} \prod_{q\in q^*} \exp\left\langle V_{c;t|t-1}, C(x_{q;t})\right\rangle + \alpha_{0;t|t-1}\mathcal{U}(x(q^*)). \tag{19}$$

The finite number of components $\mathring{c}_{t|t-1}+1$ may vary with time. The probabilistic weights $\alpha_{c;t|t-1}$ sum to unity. The values of statistics $V_{c;t|t-1}$ characterize individual components and specify their practical domain on which they are higher than the flat pdf describing unlearnt background of physical states.

Application of the Bayes rule to this prior pdf with the mixture models of sensors (18) gives, see. (6):

$$f(x(q^*)|\mathcal{P}_t) \propto \prod_{s\in s^*}\left[\alpha_s \prod_{q\in q^*} \exp\left\langle B_s(\Psi_{s;t}), H_s C(x_{q;t})\right\rangle + (1-\alpha_s)\mathcal{U}(x(q^*))\right] \times \tag{20}$$

$$\times \left[\sum_{c=1}^{\mathring{c}_{t|t-1}} \alpha_{c;t|t-1} \prod_{q\in q^*} \exp\left\langle V_{c;t|t-1}, C(x_{q;t})\right\rangle + \alpha_{0;t|t-1}\mathcal{U}(x(q^*))\right] \equiv$$

$$\equiv \sum_{c=1}^{\tilde{\mathring{c}}_{t|t}} \tilde{\alpha}_{c;t|t} \prod_{q\in q^*} \exp\left\langle \tilde{V}_{c;t|t}, C(x_{q;t})\right\rangle + \tilde{\alpha}_{0;t|t}\mathcal{U}(x(q^*)). \tag{21}$$

Thus, for the uniform background, the mixture form (19)) reproduces. The algebraic updating of statistics:

$$\left\{\alpha_{c;t|t-1}, V_{c;t|t-1}\right\}_{c=1}^{\mathring{c}_{t|t-1}} \underbrace{\rightarrow}_{\{B_s(\Psi_{s;t},H_s,\alpha_s\}_{s\in s^*}\}} \left\{\tilde{\alpha}_{c;t|t}, \tilde{V}_{c;t|t}\right\}_{c=1}^{\tilde{\mathring{c}}_{t|t}}$$

is just needed to perform the exact data updating of the global estimate. However, the number of terms $\tilde{\mathring{c}}_{t|t}$ is much higher than $\mathring{c}_{t|t-1}$ and the growth of the number of components must be limited by a suitable projection:

$$\tilde{\alpha}_{0;t|t},\ \left\{\tilde{\alpha}_{c;t|t}, \tilde{V}_{c;t|t}\right\}_{c=1}^{\tilde{\mathring{c}}_{t|t}} \rightarrow \alpha_{0;t|t},\ \left\{\alpha_{c;t|t}, V_{c;t|t}\right\}_{c=1}^{\mathring{c}_{t|t}} \quad \text{with } \mathring{c}_{t|t} = \mathring{c}_{t|t-1}.$$

There is a range of ways how to construct such a projection. The following one seems to be adequate.

Let us denote $\tilde{f}_{t|t}(x(q^*))$ the exact posterior pdf and $f_{t|t}(x(q^*))$ the constructed projection. We would like to select this projection as the minimizer of the Kullback-Leibler (KL) divergence $\mathcal{D}(\tilde{f}_{t|t}||f_{t|t}) \equiv \int \tilde{f}_{t|t}(\bullet)\ln\left(\tilde{f}_{t|t}(\bullet)/f_{t|t}(\bullet)\right) d\bullet$ (Kullback & Leibler 1951) that is known to be a good measure of proximity of pdfs. The choice is motivated by the specific role of this divergence in the considered Bayesian framework (Berec & Kárný 1997). The mixture forms of , $\tilde{f}_{t|t}, f_{t|t}$ imply that this projection cannot be practically found. Instead of it, we use the upper bound on the KL divergence implied by Jensen inequality. Its minimizer is, however, a model with a single component. In order to prevent this degeneracy, we also require proximity of the weights $\alpha_{t|t-1}$ and $\alpha_{t|t}$. Thus, we construct $\alpha_{0;t|t},\ \left\{\alpha_{c;t|t}, V_{c;t|t}\right\}_{c=1}^{\mathring{c}_{t|t}}$ as minimizing argument of:

$$-\sum_{c=1}^{\mathring{c}_{t|t}}\left[\alpha_{c;t|t}\left\langle V_{c;t|t}, \int \tilde{f}(x(q^*))\sum_{q\in q^*} C(x_q)\,dx(q^*)\right\rangle - \Lambda\alpha_{c;t|t-1}\ln(\alpha_{c;t|t})\right] + \alpha_{0;t|t} - \Lambda\alpha_{0;t|t-1}\ln(\alpha_{0;t|t}). \tag{22}$$

The optional weight $\Lambda > 0$ defines the degree of conservatism with respect to similarity of $\alpha_{t|t-1}$ and $\alpha_{t|t}$.

3.8 Time Updating of the Global Estimate

As recalled above, the time updating of the global pdf $f(x_t(q^*)|\mathcal{P}_t) \rightarrow f(x_{t+1}(q^*)|\mathcal{P}_t)$ requires unavailable time evolution model $f(x_{t+1}(q^*)|\mathcal{P}_t, x_t(q^*))$. This, in conjunction with the expected intensive informational flow from sensors, makes us to apply stabilized forgetting (Kulhavý & Zarrop 1993) on $f_{t|t}(\cdot)$. Again, mixture

form of the model prevent us to use it directly. Instead of it, we apply this forgetting component-wise. It gives the final correction of statistics

$$V_{c;t+1|t} = \lambda V_{c;t|t} + (1-\lambda)V_A.$$

The optional $\lambda \in (0,1)$ can be interpreted as the probability of the hypothesis that $x_t(q^*)$ remained constant between two consecutive measurements. The externally supplied statistic V_A describes our belief into moves of the $x_t(q^*)$ expected in the same time interval. Typically, V_A is recommended to coincide with the statistics describing the flat pdf. For bounded q^*, the flat pdf can be uniform and $V_A = 0$.

3.9 Overall Algorithm

Initial phase

- Specify the statistic V_A describing flat prior pdf $\mathcal{U}(x(q^*))$.
- Set the number of component $\mathring{c}_{1|0} = 1$ in mixture describing prior global estimate of physical state $f(x_1(q^*)|\mathcal{P}_0) \propto 1$ and initialize the corresponding statistic $V_{1;1|0} = V_A$.
- Specify structure of models of used sensors (18), i.e. their forms including selectors H_s, regression vectors $\psi_{s;t}$ and weights α_s defining mixture form of sensor models.
- Fill initial values into regression vectors of respective sensors.
- Choose the optional scalar $\Lambda > 0$ driving the mixture projection.
- Specify the upper bound $\mathring{c}$ on the number of components $\mathring{c}_{t|t}$ of the mixture describing the global estimate $f(x_t(q^*)|\mathcal{P}_{t-1})$.
- Choose the forgetting factor $\lambda \in (0,1)$ reflecting the expected rate of changes of physical states $x_t(q^*)$.

On line phase running for $t \in t^*$

- Apply external inputs $u_{s;t}$ to respective sensors (if present) demanding, for instance, to turn direction of sensing to the hottest point in space found up to now.
- Fix positions of sensors.
- Complete specification of those sensor-model characteristics that depend on their current position.
- Measure sensors outputs $y_{s;t}$ and complete data vectors $\Psi_{s;t} = [y_{s;t}, \psi_{s;t}]$.

Data updating % Fusion of Sensor Readings

- Update statistics $V_{c;t|t-1}$, $c = 1, \ldots, \mathring{c}_{t|t-1}$ to $\tilde{V}_{c;t|t}$, $c = 1, \ldots, \tilde{\mathring{c}}_{t|t}$ using values of functions $B_s(\Psi_{s;t})$, H_s.
- Update weights $\alpha_{c;t|t-1}$, $c = 0, 1, \ldots, \mathring{c}_{t|t-1}$ to $\tilde{\alpha}_{c;t|t}$, $c = 0, 1, \ldots, \tilde{\mathring{c}}_{t|t}$ using values of functions $B_s(\Psi_{s;t})$, H_s, α_s and normalization to their unit sum.

Projection

- Set $V_{c;t|t} = \tilde{V}_{c;t|t}$, $\alpha_{c;t|t} = \tilde{\alpha}_{c;t|t}$, $\mathring{c}_{t|t} = \tilde{\mathring{c}}_{t|t}$ if $\tilde{\mathring{c}}_{t|t} \leq \mathring{c}$ and go to **Time updating by forgetting**.
- Project $\tilde{V}_{c;t|t}$, $\tilde{\alpha}_{c;t|t}$ on $V_{c;t|t}$, $\alpha_{c;t|t}$ by minimization of (22) if $\tilde{\mathring{c}}_{t|t} > \mathring{c}$ and continue.

Time updating by forgetting

- Forget by defining $V_{c;t+1|t} = \lambda V_{c;t|t} + (1-\lambda)V_A$.
- Evaluate point estimates of state, if need be, and go to the beginning of On line phase.

Remarks

1. Extent of prior options determines to significant extent usefulness of algorithm. Let us discuss them
   (a) The choice of statistics V_A describing flat background and prior pdf is inevitable and relatively simple. Properties of the algorithm are expected to be reasonably insensitive to this choice.
   (b) Forgetting factor is relatively easy to choose and the algorithm will be robust to its choice.
   (c) The weighting factor $\Lambda > 0$ is introduced in heuristic way and there is no experience with its choice and influence. *Its choice can be almost surely avoided by approximating joint pdf of data and pointers to components*. Such a pdf is known as an extension that has the mixture model as its marginal. Similar "trick" can be applied when introducing forgetting.
   (d) The upper bound on the number of components $\mathring{c}$ should be chosen as high as computationally acceptable. Its value is determined by computation resources available and by desired sampling rate.
2. Computational demands of the algorithm depend on dimensionality of the problem (length of x_t, number of sensors and dimensions of their outputs and number of components in the global estimate). Preliminary

considerations and the particular case reported in (Beal & Jojic 2003) indicate that the computational load can be acceptable in a wide range of applications.
3. Updating of statistic $V_{c;t|t-1} \rightarrow \tilde{V}_{c;t|t}$ in the dominant case of normal components is equivalent to single step of recursive least squares (Peterka 1981). As such it is computationally relatively cheap.
4. The considered projection performs merging and cancelling components at one shot. Its computational complexity is expected to be around that needed for data updating. Simpler versions of projections are at disposal if this load will be found to high.

4 Robot Platform Considered for Tests

A special robot head has been constructed as a test bed for multi-sensor integration and convergence. The head has been positioned as a hand on the end of a 6 Degrees of Freedom Industrial Robot Manipulator Arm. As a result the head can take up any position and attitude in respect to an object in 3 dimensional space.

The robot head for experimentation (see Figure 2) has been named MORGUI, which is Mandarin Chinese for Magic (mor) Ghost (gui). It consists of a rapid action head containing 5-senses. Two of these, vision and audio, are human equivalents, whilst the remainder radar, infrared and ultrasonics are extrasensory as far as humans are concerned.

Morgui's camera vision can be either fixed stereo (i.e. the eyes cannot be separately rotated) or mono, with one camera being employed for sensory purposes within the robot, whilst the other is used for monitoring, observation and recording of (possibly human) object responses. Audio meanwhile consists of two receivers positioned on either side of Morgui's head, positioned a little forward from the possibly expected (human ear) location in order to reduce noise effects at central positions.

Ultrasonic sensors located in Morgui's forehead give a broad object appearance signal at a fairly accurate range, i.e. the sensor output give a reasonably accurate ($\pm$ 2%) indication of the distance to the nearest object of a reasonable size. In Morgui's nose is a Doppler radar system, which gives an accurate indication of the speed of movement of an object, to or from the robot head. Finally, positioned just above Morgui's top lip is a dual infrared sensory system, which can give a reasonable thermionic read out from a specified distance away or, more likely, can be used to give an alternative distance measure from the robot head to the hottest nearby object. In the sense of a distance measure the ultrasonic and infrared sensors can be deemed to be mutually corresponding in that they are indicating different views on the same measurement – distance from the robot head. Further views on the same measure can also be obtained through a depth calculation using the stereovision system, if the magnification of object audio output is known and if the object moves, through radar. For our own studies we will initially consider input on distance from the robot head through signals taken in via the ultrasonic and infrared sensors.

The Control problem with Morgui is one that with the robot being given a straightforward single goal, it is this goal only that it aims to achieve at all times. Decision making between multiple goals through competing actors for control of the robot (as in (Breazeal 2000)) is not the subject of study here. The critical factor in our case is multisensory correspondence, enabling the robot to achieve its single goal with the aid of diverse sensory information.

A general directive for Morgui could be to retain its head position in uniform 3 dimensional space with regard to a distant object, i.e. to always retain a specified distant object in the center of the robot's sensory system, with the object remaining a set, pre-chosen, distance away. The aim being that Morgui focuses on, and tracks in real-time, a distant, specific moving objects. To restrict the goal, in the first instance to something a little simpler – we can merely give Morgui the task of keeping the distance, in the z-direction, between itself and an object, constant with respect to time. The aim is therefore for Morgui's head position to remain a set distance from an object as it moves in z-space only.

It is worth noting that the control algorithm employed with Morgui, for target tracking, is not of major concern here, in fact a PI control will invariably suffice. The only requirement is for the algorithm to ensure that the robot responds in a reasonable time frame (with regard to the object's speed) and with reasonable accuracy. Our concern in foreseen tests is in essence not about the type of control to be employed, but rather that whichever controller it is, the accuracy of the information it has at its disposal is improved through the use of multisensory convergence.

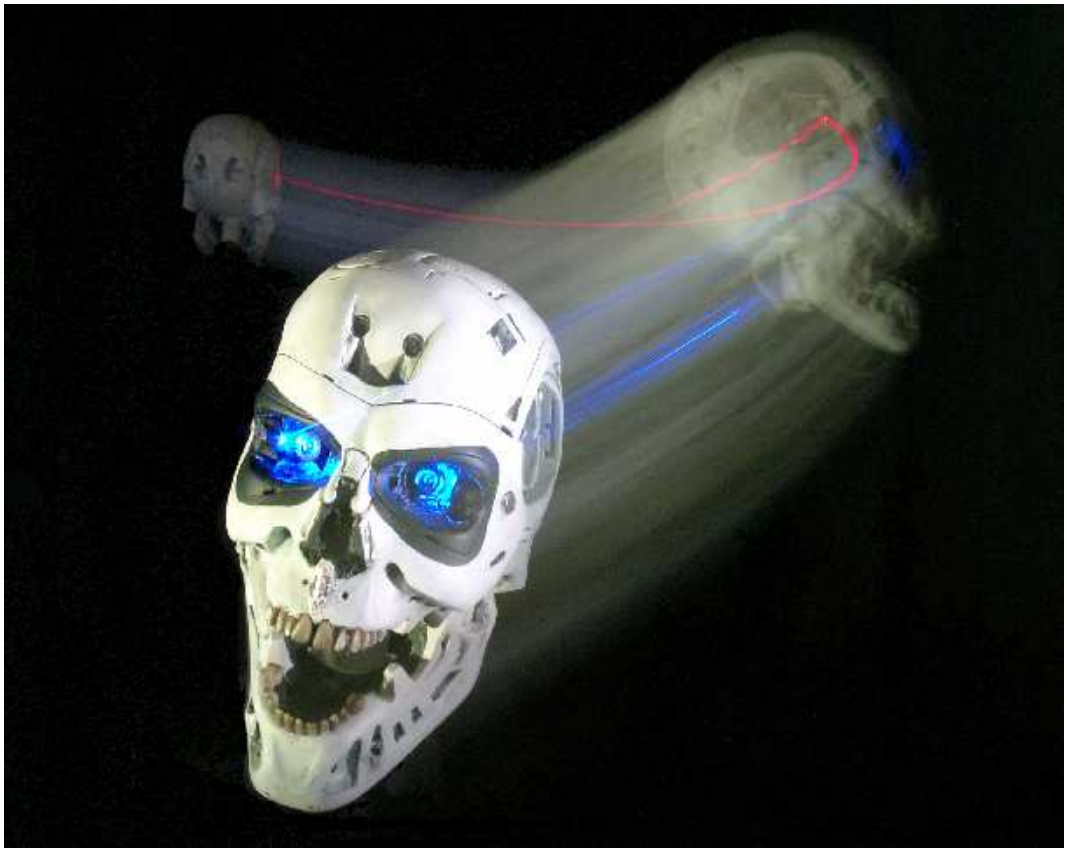

Figure 2. Morgui – Multisensory Robot Head in Action

5 Concluding Remarks

This preliminary inspection of the general but applicable framework of sensor fusion has been started from two distinctly different schools of research, namely state-space analysis of sensors as dynamic system and stochastic filtering reduced to Bayesian estimation of mixtures combined with advanced forgetting techniques. The paper presents a promising but still preliminary synthesis of this attempt to tackle the general sensor fusion problem. The current state of the research outlines further steps:

- Modelling of Morgui's sensors in the form described in section 3.5.
- Elaboration of algorithmic details of global fusion algorithm summarized in section 3.9.
- Implementation of the algorithm and extensive tests on the robot head test bed that should help to refine and if need be to modify the described approach.
- Modification and possible adaptation of the theory behind to the discussed area.

This is an ambitious plan but the framework presented in the paper promises that it is feasible. The generic nature of the solution indicates that it will be applicable not only to Murgui's head.

Acknowledgments

This work was supported by ESF, project TED, GAČR 102/03/P010 and GAČR 102/03/0049.

References

B. Anderson & J. Moore (1979). *Optimal Filtering*. Prentice Hall.
O. Barndorff-Nielsen (1978). *Information and exponential families in statistical theory*. Wiley, New York.

M. Beal & N. Jojic (2003). 'A graphical model for audiovisual object tracking'. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **25**(7):828–836.

L. Berec & M. Kárný (1997). 'Identification of reality in Bayesian context'. In K. Warwick & M. Kárný (eds.), *Computer-Intensive Methods in Control and Signal Processing: Curse of Dimensionality*, pp. 181–193. Birkhäuser.

L. Berthouse & Y. Kunniyoshi (1998). 'Emergence and categorization of coordinated visual behaviour through embodied interaction'. *Autonomous Robots* **5**(3–4):369–379.

T. Boss, et al. (2001). 'Sensor fusion by neural networks using spatially represented information'. *Biological Cybernetics* **85**:371–385.

C. Breazeal (2000). *Sociable Machines, Expressive Social Exchange Between Humans and Robots*. Ph.D. thesis.

E. Daum (1988). 'New exact nonlinear filters'. In J. Spall (ed.), *Bayesian Analysis of Time Series and Dynamic Models*. Marcel Dekker, New York.

C. Engels & G. Schoner (1995). 'Dynamic fields endow behaviour based robots with representations'. *Journal of Robotics and Autonomous Systems* **14**:55–77.

G. Foresti & C. Regazzoni (2002). 'Multisensor data fusion for autonomous vehicle navigation in risky environments'. *IEEE Transactions on Vehicle Technology* **51**(5):1165–1185.

I. Howard (1998). 'Interactions within and between the spatial senses'. *J. Vestib. Research* **7**:311–345.

R. Jassemi-Zourgani & D. Necsulescu (2002). 'Extended Kalman filter based sensor fusion for operational space control of a robot arm'. *IEEE Transactions on Instrumentation and Measurement* **51**(6):1279–1282.

A. Jazwinski (1970). *Stochastic Processes and Filtering Theory*. Academic Press, New York.

R. Kulhavý & M. B. Zarrop (1993). 'On general concept of forgetting'. *International Journal of Control* **58**(4):905–924.

S. Kullback & R. Leibler (1951). 'On information and sufficiency'. *Annals of Mathematical Statistics* **22**:79–87.

G. McLachlan & D. Peel (2001). *Finite Mixture Models*. Wiley, New York.

M. Meredith (2002). 'On the neuronal basis for multisensory convergence: a brief overview'. *Cognitive Brain Research* **14**:31–40.

V. Peterka (1981). 'Bayesian system identification'. In P. Eykhoff (ed.), *Trends and Progress in System Identification*, pp. 239–304. Pergamon Press, Oxford.

A. Quinn, et al. (2003). 'On applications of mixture models'. *Int. J. of Adaptive Control and Signal Processing* **17**:133–148.

D. Robinson (1977). 'Linear Addition of Optokinetic and Vestibular signals in the Vestibular Nucleus'. *Experimental Brain Research* **30**:447–450.

B. Stein & M. A. Meredith (1993). *The merging of the senses*. MIT Press, Cambridge MA.

D. Titterington, et al. (1985). *Statistical Analysis of Finite Mixtures*. Wiley, New York.

T. Tyrrell (1994). 'An evaluation of Maes's Bottom up Mechanism for Behaviour Selection'. *Adaptive Behaviour* **2**(4):307–348.

R. van Beers, et al. (99). 'Integration of proprio ceptive and visual position information: an experimentally' .

K. Warwick (1987). 'Optimal Observers for ARMA models'. *Int. J. Control* **46**(5):1493–1503.

J. Wolfe (1994). 'Guided Search 2.0: A revised model of visual search'. *Psychonomic Bulletin and Review* **1**(2):202–238.