2013 IEEE/RSJ International Conference on
Intelligent Robots and Systems (IROS)
November 3-7, 2013. Tokyo, Japan

Learning to combine multi-sensor information for context dependent state estimation

Alexandre Ravet[1], Simon Lacroix[1], Gautier Hattenberger[2] and Bertrand Vandeportaele[1]

Abstract— **The fusion of multi-sensor information for state estimation is a well studied problem in robotics. However, the classical methods may fail to take into account the measurements validity, therefore ruining the benefits of sensor redundancy. This work addresses this problem by learning context-dependent knowledge about sensor reliability. This knowledge is later used as a decision rule in the fusion task in order to dynamically select the most appropriate subset of sensors. For this purpose we use the Mixture of Experts framework. In our application, each expert is a Kalman filter fed by a subset of sensors, and a gating network serves as a mediator between individual filters, basing its decision on sensor inputs and contextual information to reason about the operation context. The performance of this model is evaluated for altitude estimation of a UAV.**

I. Introduction

State estimation is an essential issue in robotics. For many systems, it relies on multiple sensors, each one exhibiting an inherent observation uncertainty, operating range, and context dependent performance. Uncertainty due to the observation noise received considerable attention over past decades and is commonly handled using Bayesian filtering [1]. It is also well known that the use of redundant sensors significantly improves estimation accuracy and reliability. However, such methods do not provide any satisfying way to assess the validity of sensor measurements.

In the context of multi-sensor state estimation, most attempts to deal with this issue lead to self-contained systems, relying on information theoretic framework [2] or rejection schemes designed after experience on the system behaviour [3]. This work is motivated by the fact that an intelligent system should not only be able to select the sensor -or subset of sensors- based on an online performance measure, but should also encode knowledge about the reliability of a perception modality according to the current specific context.

This implies the ability for the system to discover the implicit operation contexts the robot is likely to encounter, based on the a priori unknown performances of each sensor in these contexts. Attributing belief about the reliability of a sensor in these contexts then requires a complex reasoning on information acquired about the environment. Except for simple cases (reduced set of sensors, known environment) one can not easily implement such decision rules by hand.

[1]CNRS; LAAS; 7 avenue du colonel Roche, F-31077 Toulouse, France ; Université de Toulouse ; UPS, INSA, INP, ISAE; LAAS; F-31077 Toulouse, France ; (aravet), (simon.lacroix), (bvandepo) at laas.fr

[2]Ecole Nationale de l'Aviation Civile (ENAC), 7 Avenue Edouard Belin, BP-54005, Toulouse, France. gautier.hattenberger at enac.fr

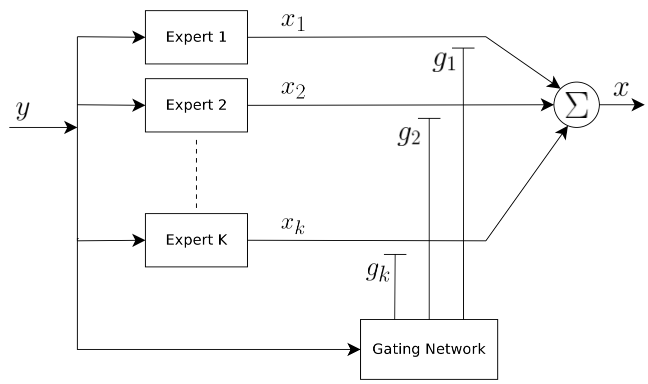


Fig. 1. The basic mixture of experts framework

Addressing this problem introduces the need for the system to learn how to achieve the sensor selection task. For this purpose, we propose to use a supervised learning algorithm to learn a mapping from sensors measurements input space (and any relevant information) to sensors reliability probabilities.

A well-designed robotic platform should exhibit various perception modalities relying on different but complementary physical principles. Consequently the set of perception modalities embedded on a robot does not provide direct commensurate measurements, and it is often easier and more modular to fuse information at a state vector level [4]. Furthermore, binding different subset of sensors to different estimation filters allows to map the sensor selection problem to the bank of Kalman filter approach. This method assumes that optimal filtering can be expressed by dynamically selecting the most suitable filter among a bank of filters. This approach emerged with the Magill's filter bank [5], and has been subsequently improved leading to general pseudo Bayes (GPB) methods and interacting multiple models (IMM) [6], the latter being more computationally efficient. Although some authors decided to augment the IMM with context-dependent information [7], this algorithm fundamentally relies on the exploitation of internal estimates and a known transition probability matrix between different filter models. Thus introduction of a knowledge about context dependent model reliability is not straightforward, especially if the user wants the system to learn this information. An analogous approach can be found in [8] in the context of fault diagnosis. Based on a jump Markov linear model, this method requires a priori knowledge of the different regimes of operation for the learning step, while we want our system to discover these different contexts by itself.

978-1-4673-6357-0/13/$31.00 ©2013 IEEE

Aiming at learning how to combine some complementary experts, the Mixture of Experts (ME) framework lends itself very well to the problem as it basically computes an optimal output through a weighted sum of individual experts. To achieve this mediation task, the ME relies on a gating network in charge of providing gating probabilities, equivalent to reliability coefficients over the set of experts (Fig. 1). When experts are replaced by estimation filters, this approach is known to be an efficient alternative to the filter bank approach [9] [10].

This article aims at showing how the mixture of Kalman filter for implicit sensor selection can be applied to the altitude estimation task for a UAV, and is based on the two following contributions :

- Application of the localized gating network to the mixture of Kalman filters
- Application of the bank of Kalman filter approach for implicit sensor selection

It is organised as follows: Section II introduces the ME framework and concept of adaptive Kalman filtering for sensor selection. Section III focuses on the gating model and the training phase of the ME. Section IV conveys the experimental results obtained for simulation and real data scenario. Concluding remarks are finally made in section V.

II. Theoretical Background

A. The mixture of experts framework

The mixture of experts approach basically consists in decomposing a complex problem into subtasks, each of which being handled by an appropriate expert. Traditionally used for regression or classification problems, the model learns to split the input space into overlapping regions within which assigned experts are active.

The standard ME framework [11] consists in a set of K experts modules and a gating network (Fig. 1). Each expert $k = 1...K$ associated with parameters λ_k looks at input vector y and computes a local output x_k through a function $f_k(\lambda_k, y)$. In a probabilistic interpretation, the output of an expert k can be viewed as the mean of a probability distribution $P(x|y, \lambda_k)$ with x the desired target value associated to sample y. Assuming that the different experts may be more competent in different regions of the input space (*i.e.* they have higher probability to produce the desired target x), the gating network mediates the outputs of the bank of experts by producing for each expert k a probability of its output x_k to be equal to the desired output x. This results in a set of gating probabilities g_k weighting the output of all experts while satisfying constraints $g_k \geq 0, k = 1...K$, and $\sum_{k=1}^{K} g_k = 1$.

Given an input vector y and a target vector x, the probability of observing x is then written in terms of gating probabilities and experts outputs (using product rule) as

$$\begin{aligned} P(x|y, \Theta, \Lambda) &= \sum_{k=1}^{K} P(x, k|y, \Theta, \Lambda) \\ &= \sum_{k=1}^{K} P(k|y, \Theta) P(x|k, y, \Lambda) \\ &= \sum_{k=1}^{K} g_k(y, \theta_k) P(x|y, \lambda_k) \end{aligned} \tag{1}$$

where $\{\Theta, \Lambda\}$ denotes the set of all parameters, with $\Theta = \{\theta_k, k = 1...K\}$ the set of gate parameters and $\Lambda = \{\lambda_k, k = 1...K\}$ the set of experts parameters.

ME implementations then differ in three main points: the experts model, the gating model, and the inference method [12]. Our model for the gating framework is justified in section II, and expert models are set as Kalman filters in our case. In this paper we use a common learning method based on the maximum likelihood principle, quickly described hereafter.

Given a training set $\{\mathbf{x}, \mathbf{y}\}$ we try to maximize the likelihood $\mathcal{L}$ of the data set with respect to the model parameters. If samples are considered identically independently distributed, this is equivalent to maximize:

$$\mathcal{L} = \prod_{n} p(x^n, y^n)$$

We then define the usual cost function $\mathcal{C}$ as the negative log of the likelihood function, such that maximizing likelihood is now equivalent to minimize $\mathcal{C}$:

$$\mathcal{C} = -\sum_{n} ln(p(x^n, y^n))$$

Different methods for determining max likelihood have been developed. The standard gradient descent methods can be applied. More recently, sampling, variational inference and several Expectation Maximization (EM) algorithms have emerged [12] and have shown good performances.

B. Adaptive Kalman filtering for sensor selection

We implicitly solve the sensor selection problem through the filter bank approach. In classical implementations, the bank is composed of a finite number of filters differing in transition model, transition noise and observation noise. A weighting function then assigns weight factors to the output of each individual filter, giving highest weight value to the best performing filter. In our case, models differ only in observation matrices (and corresponding observation noise), acting as a selector on the sensors (Fig. 2). By this mean we implicitly select the most appropriate subset of sensors through the filter selection process.

III. ME framework for sensor selection

A. Motivations

A micro-UAV is often brought to deal with changing environment. The simple take-off and landing phase of a micro-UAV already brings many different regimes in term

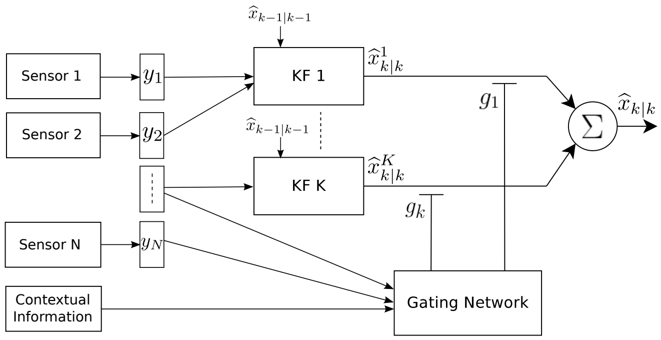


Fig. 2. Mixture of experts framework for sensor selection. Each filter can be wired to a different subset of sensors. The gating network can share experts inputs and use any useful additional contextual information.

of sensor performance for altitude estimation. Ultrasonic sensors are for example quite reliable and accurate until they reach a given maximum range. They also easily provide outliers measurements, *e.g.* because of multiple reflections. Vision may start to provide information after reaching a given altitude, depending on camera characteristics and on the ground texture. Barometric pressure sensors provide wide measuring range with quite constant accuracy but also require to estimate a bias due to changing atmospheric conditions, while GPS provides signal dependant precision, and is more likely to be reliable for high altitudes, also depending on environment characteristics. These specificities raise the need to create decision rules for selecting an active sensor subset given a specific context.

In [3] the authors report for example that indoor/outdoor transitions result in outliers classical methods can not reject. Therefore, a mechanism for sensor selection is proposed, giving ability to the system to switch to the sensor that works well in the current environment. The selection rule relies on the strong assumption that the sensor with the smallest measurement variance is the more reliable. If this approach turns to be efficient in this specific transition context and sensor setup, it is not generic at all.

An important capability of the ME model, which motivated this work, relies in the gating network ability either to share the experts inputs or to use additional information. This allows to base decision on context-dependent information of any kind. Under the assumption that the training set contains enough samples, the learned gating network then ensures adaptation to the different environments, providing a partial assessment of sensors reliability.

B. Using Localized gating network to encode decision rules

Besides adapting the perception modalities, we also aim at switching smoothly between experts. This requirement especially makes sense in flight context, where hard transitions between sensors (consequently between estimates) are not admissible, as it directly impacts the robot safety in cluttered environments. In the standard model, the gating network is a single layer linear network, hence the decision boundaries consist of 'soft' hyperplanes and inevitably create overlapping regions [13], within which only one expert may be needed (*i.e.* only one sensor subset is effective). Consequently we adopt a specific model for the gating network, known as localized ME [14], which consists of normalized Gaussian kernels (or any density function from the exponential family) :

$$g_k(y, \theta_k) = P(k|y) = \frac{\alpha_k P(y|\theta_k)}{\sum_{j=1}^{K} \alpha_j P(y|\theta_j)} \tag{2}$$

with

$$P(y|\theta_k) = \frac{1}{(2\pi)^{d/2} | \Sigma_k |^{1/2}} exp\left(-\frac{(y-m_k)^T \Sigma_k^{-1} (y-m_k)}{2}\right)$$

where $\theta_k = \{m_k, \Sigma_k\}$ the mean and variance of the Gaussian kernel distribution.

The Gaussian kernels allow to divide the input space into soft hyper-ellipsoids. These ellipsoids can overlap, or create localized regions of expertise where a single sensor is trustworthy. The choice of Gaussian kernels also impacts the learning step, as it yields a one-pass maximization step when using the EM algorithm. EM is proved to have a faster convergence rate than gradient ascent methods [15], and provides guaranteed convergence due to the single loop maximization step when used with Gaussian kernels: hence we learn the gating parameters with the EM algorithm.

C. Learning the mixture parameters

The basic idea of the EM algorithm is to make the assumption that some variables are hidden, in our case the probability that the n^{th} target sample x^n was generated by expert k. Hence we introduce an indicator variable z :

$$z_j^n = \begin{cases} 1 & \text{if target sample } x^n \text{ is generated by expert } j \\ 0 & \text{otherwise} \end{cases}$$

This hidden variable induces mutual competition among experts. It also models the existence of unknown operating contexts which for different subsets of experts are reliable. To obtain a one pass calculation for the gating parameters, we complete maximum likelihood estimation on the joint density $p(x, y)$ [14]. Rewriting equation (1) with the new gating function and noting the k^{th} expert output conditional density function $\phi_k(x|y)$:

$$p(x|y, \Theta, \Lambda) = \sum_{k=1}^{K} \frac{\alpha_k P(y|\theta_k)}{\sum_{j=1}^{K} \alpha_j P(y|\theta_j)} \phi_k(x|y) \tag{3}$$

we obtain the joint density

$$p(x, y) = \sum_{k=1}^{K} \alpha_k P(y|\theta_k) \phi_k(x|y) \tag{4}$$

using Baye's rule on (2) to obtain $p(y) = \sum_{j=1}^{K} \alpha_j P(y|\theta_j)$.

Finally, introducing the indicator variable z to mediate mutually exclusive experts, the joint distribution over hidden and observed variables takes the form :

$$p(x, y, z) = \prod_{k=1}^{K} (\alpha_k P(y|\theta_k) \phi_k(x|y))^{z_k} \tag{5}$$

which by maximum likelihood leads to the cost function:

$$\mathcal{C} = -\sum_n \sum_{k=1}^{K} z_k^n ln(\alpha_k P(y^n|\theta_k)\phi_k(x^n|y^n)) \quad (6)$$

Now the specificity of EM algorithm enters. In the expectation step we replace the hidden variable z by its expected value :

$$\begin{aligned}\mathcal{E}(z_k^n) &:= p(z_k^n = 1|x^n, y^n) \\ &= \frac{p(x^n|z_k^n = 1, y^n)p(z_k^n = 1|y^n)}{p(x^n|y^n)} \\ &= \frac{\alpha_k P(y^n|\theta_k)\phi_k(x^n|y^n)}{\sum_{j=1}^{K} \alpha_j P(y^n|\theta_j)\phi_j(x^n|y^n)} = h_k(x^n, y^n) \quad (7)\end{aligned}$$

Then the maximization step maximizes the expectation of the cost function by substituting z_k by its expectation $h_k(x, y)$.

$$\mathcal{E}(\mathcal{C}) = -\sum_n \sum_{k=1}^{K} h_k(x^n, y^n) ln(\alpha_k P(y^n|\theta_k)\phi_k(x^n|y^n)) \quad (8)$$

As we can see this cost function can be separated in two terms. The first one corresponds to the cost function relative to gating parameters ($\alpha_k P(y^n|\theta_k)$) and the second term corresponds to the expert network parameters.

D. Achieving mixture of Kalman filter

In our context each expert is a particular Kalman filter providing its own estimation based on observation input y_k^n and parameters λ_k describing specific sensor observation noise and observation selection matrix. Hence $\phi_k(x^n|y^n)$ is obtained by evaluating the output distribution of the k^{th} Kalman filter at point x^n. The maximization step then consists only in minimizing the first term of result (8). Setting partial derivatives w.r.t to α_k (and using Lagrangian multiplier to introduce the constraint $\sum_k \alpha_k = 1$), m_k and Σ_k to zero, we obtain new estimates [13]:

$$\alpha_k = \frac{1}{N} \sum_n h_k(x^n, y^n) \quad (9)$$

$$m_k = \frac{\sum_n h_k(x^n, y^n) y^n}{\sum_n h_k(x^n, y^n)} \quad (10)$$

$$\Sigma_k = \frac{1}{d} \frac{\sum_n h_k(x^n, y^n) \| y^n - m_k \|^2}{\sum_n h(x^n, y^n)} \quad (11)$$

Using these new parameters, we then repeat the EM steps until convergence.

One common problem with mixture of Kalman filter is that the exact belief state grows exponentially in time. For a set of K filters, at iteration $t = T$, the exact distribution of the state is a mixture of K^T Gaussian distributions. To deal with this exponential growth we use the GPB collapsing method of order 1, and approximate the mixture of filters output distribution with a single Gaussian distribution. At step n, if each filter k provides an output distribution of mean μ_k and variance σ_k we obtain the mixture distribution mean μ_{mix} and variance σ_{mix} [16]:

$$\mu_{mix} = \sum_{k=1}^{K} g_k \mu_k$$

$$\sigma_{mix} = \sum_{k=1}^{K} g_k [\sigma_k + (\mu_k - \mu_{mix})(\mu_k - \mu_{mix})^T]$$

The next transition step is then based on this mixture output, hence accumulating the error introduced by the approximation at each time step. However, it has been shown in [17] that the process error remains bounded indefinitely, avoiding the mixture output to become irrelevant.

Some drawbacks of the approach are now discussed. In its original implementation the ME framework inputs are synchronized, and the gating network bases its decision on a joint set of observations. For experiments, we simulated synchronous observations by forcing sensors to provide measures at a defined frequency. As we will see, this approach does not affect the framework ability to make decisions, mainly because difference between inputs frequencies are small – however large differences would not provide relevant decisional capabilities.

An other constraint, directly imposed by the Gaussian kernel model, is the unimodal distribution of the regions of expertise. Under specific configurations, some sensors may need to be active in separate regions of the input space. This would require to model gating probabilities with more complex models, like Gaussian mixtures or Gaussian processes [12] – we will however notice that in our application context the localized Gaussian kernels provide good behaviour.

IV. EXPERIMENTS

A. Simulation

We first illustrate the system ability to learn decision rules according to sensors characteristics. This simple example reproduces the take-off and landing phases of a UAV. Three sensors provide direct measures of the altitude with different characteristics, such as observation noise, outliers occurrences and measurement range thresholds (Fig. 3). Each sensor is fed to one filter, and all filters share a common constant velocity transition model. We train the gating network on a dataset of 12000 samples reproducing two subsequent take-off/landing sequences. The EM algorithm takes 50 iterations to converge with a convergence threshold of 10^{-5}.

The final estimate and associated uncertainty boundaries for the validation set is shown in Fig. 5(a). As we can see, the gating network learned to switch between sensors in order to reject outliers and to take into consideration each sensor measurement range. As expected from mutual competition between experts introduced during the learning step, the gating network tends to assign binary weights. Hence mixing only operates during transition phases. As a consequence, the system output provides consistent estimation but does not benefit from estimation uncertainty reduction that could be provided by direct measure fusion.

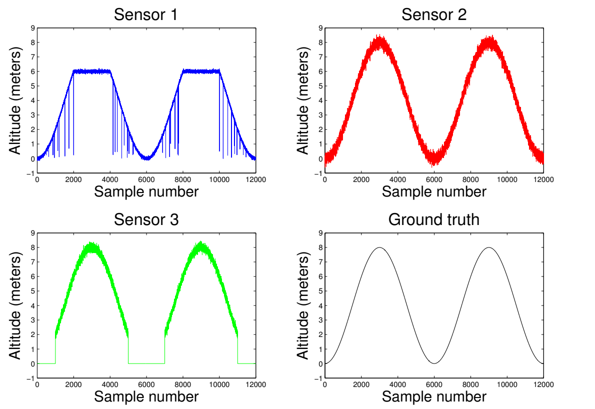


Fig. 3. Altitude measurement provided by three sensors. Sensor 1 reproduces typical ultrasonic measures, low observation noise, outliers occurrences and maximum range threshold. Sensor 2 permanently provides measures with high observation noise. Sensor 3 does not provide relevant measures below 2 meters.

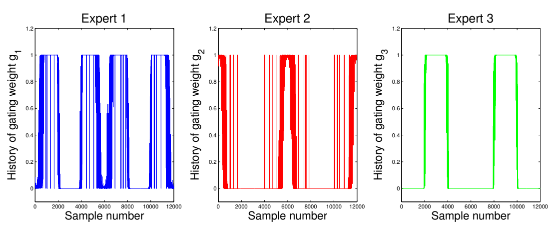
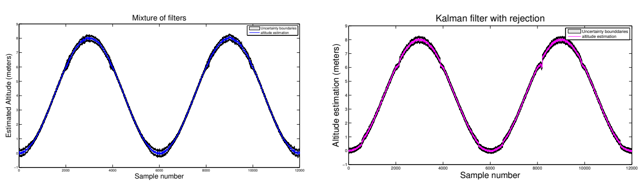


Fig. 4. Gating weights history on validation dataset.



(a) ME approach

(b) Kalman filter with rejection

Fig. 5. Estimated altitude on validation dataset.

We compared the ME approach with a classical Kalman filter enhanced with 3-sigma rejection on all sensors. As shown Fig. 5(b), this approach can provide similar results with appropriately tuned filter parameters. However, the efficiency of such methods proves to be unsound, especially as small changes in filter parameters or rejection threshold can lead to strong divergence of the estimation output. While we observed that changes in filter parameters significantly modifies the localization of Gaussian kernels in the input space, the ME approach turns to homogeneously produce consistent output thanks to its adaptation capability.

B. Real Data

We now use datasets acquired on a paparazzi quadrotor UAV [18]. Datasets consists of $50Hz$ synchronized altitude measures provided by an ultrasonic sensor and a barometer as well as accelerations on 3 axis provided by the embedded IMU. Altitude truth is given by a motion capture system. As we can see Fig. 6, ultrasonic sensor presents strong and frequent outliers we know to be related to thrust level. We also suppose that an external filter gives us an estimation of the barometer offset.

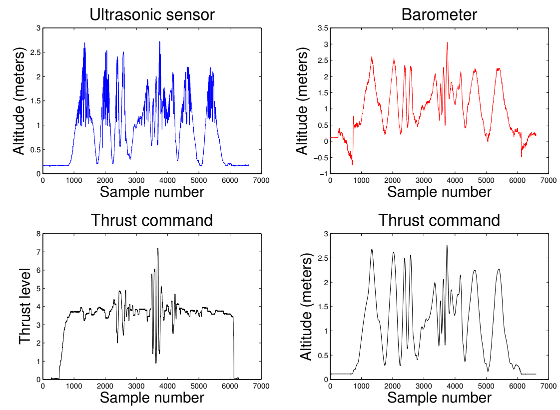


Fig. 6. Gating network inputs and altitude truth for validation dataset.

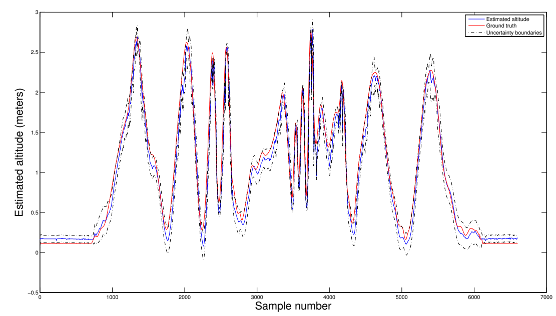


Fig. 7. Altitude estimation and uncertainty boundaries using mixture of Kalman filters on validation set.

Without additional understanding of the perturbations generated on ultrasonic sensor measures, we apply the mixture of experts framework to show its ability to learn to filter these outliers, and improve estimation accuracy. For this application we use 3 different experts: one expert based on ultrasonic measures, an other based on barometer measures, and a last one based on both ultrasonic and barometer measures. As we know the presence of outliers in ultrasonic observations is correlated to the thrust, we provide 3 inputs to the gating network: both sensor measures and the thrust command. We compare this method to a Kalman filter using 3-sigma rejection scheme on ultrasonic and barometer measures. All these filters share the same constant velocity transition model and observation noise. We train the system on a dataset of 5000 samples. After 50 iterations the EM algorithm reaches the convergence threshold fixed to 10^{-6}.

Experiments show that the learned parameters generalize well on different validation sets, always providing similar performances. As we can see Fig. 7, some outliers are not perfectly filtered. These outliers are presumably localized in unexplored regions of the input space, implying that rejection capability could be improved by using a larger training set. On the validation set corresponding to Fig .6, the system

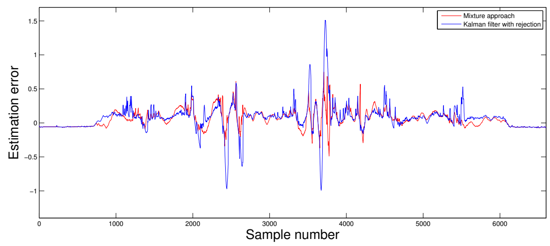


Fig. 8. Estimation error relative to altitude truth for mixture of Kalman filter (in red) and Kalman filter with 3-sigma rejection (in blue).

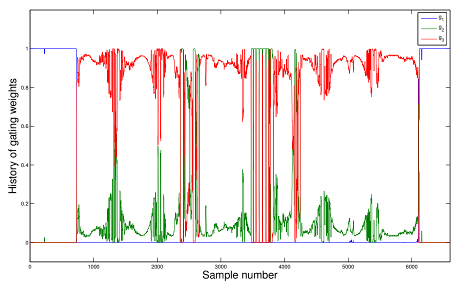


Fig. 9. Weights history for expert 1 (ultrasonic sensor + barometer), expert 2 (ultrasonic sensor) and expert 3 (barometer).

provides the best RMS estimation error with a value of 0.135. The filter with rejection provides an RMS error of 0.207. With a gating framework basing its decision on sensor measures only, we found an RMS error of 0.150. This result attests of thrust impact on ultrasonic measures, and of the ability for the framework to take it into account as well.

The estimation error improvement provided by the mixture approach (shown Fig. 8) can be explained by the sensor selection process. For example our model learned to assign more weight to the ultrasonic sensor as the UAV gets closer to the ground, and usually promotes the barometer for higher altitudes, where outliers on ultrasonic measures are more likely to appear. Note that due to its estimation latency, the barometer measures are more relevant for small velocity. This is why, based on the strong thrust command value, our approach reduces estimation error by now choosing the ultrasonic sensor during the fast transition phase between sample 3000 and 4000. The error difference provided by the Kalman filter here again results from its sensitivity on filter parameters. The noise term on the transition model should reflects the dynamics of the UAV, but as rejection is based on the innovation term, high noise terms reduce rejection capability on small outliers. At the same time, low values can conduct to correct measurement rejection during high dynamic maneuvers. This explains the highest error peaks on Fig. 8 where estimation latency introduced by the use of the barometer punctually becomes coherent with ultrasonic outliers, and make the filter diverge until altitude decreases. In this more complex example, all the sensor specificities can not be handled by an appropriate parameter tuning, and more subtle decision rules as encoded in the gating network prove to be more suitable.

V. Conclusion

We demonstrated that the mixture of expert framework can be applied to the sensor selection problem. The gating network discovers the different operating contexts and encodes knowledge about sensor reliability through the gating probability distributions parameters. This enables the system to automatically select the best suited estimation output, improving robustness regarding filter parameters inaccuracies and sensor characteristics.

An interesting direction for future work would consist in using more complex models for decision boundaries and extend the method to richer information sources like laser range data or images. In the current implementation the mixture process ignores previous gating weight values. Extending the gating network to its dynamical version would also improve performances of the approach.

References

[1] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. The MIT Press, 2005.

[2] S. R. Sukumar, H. Bozdogan, D. L. Page, A. Koschan, and M. A. Abidi, “Sensor selection using information complexity for multi-sensor mobile robot localization.” in *IEEE ICRA*, 2007, pp. 4158–4163.

[3] T. Tomic, K. Schmid, P. Lutz, A. Domel, M. Kassecker, E. Mair, I. Grixa, F. Ruess, M. Suppa, and D. Burschka, “Toward a fully autonomous uav: Research platform for indoor and outdoor urban search and rescue,” *Robotics Automation Magazine, IEEE*, vol. 19, no. 3, pp. 46 –56, sept. 2012.

[4] D. L. Hall and J. Llinas, “An introduction to multisensor data fusion,” *Proceedings of the IEEE*, vol. 85, no. 1, pp. 6–23, Jan. 1997.

[5] D. Magill, “Optimal adaptive estimation of sampled stochastic processes,” *IEEE Transactions on Automatic Control*, vol. 10, no. 4, 1965.

[6] H. Blom and Y. Bar-Shalom, “The interacting multiple model algorithm for systems with markovian switching coefficients,” *Automatic Control, IEEE Transactions on*, vol. 33, no. 8, pp. 780 –783, aug 1988.

[7] R. Schubert and G. Wanielik, “Unifying Bayesian networks and IMM filtering for improved multiple model estimation,” in *12th International Conference on Information Fusion*, 2009.

[8] N. de Freitas, R. Dearden, F. Hutter, R. Morales-Menendez, J. Mutch, and D. Poole, “Diagnosis by a waiter and a mars explorer,” *Proceedings of the IEEE*, vol. 92, no. 3, pp. 455–468, 2004.

[9] W. Chaer, R. Bishop, and J. Ghosh, “Hierarchical adaptive kalman filtering for interplanetary orbit determination,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, no. 3, 1998.

[10] W. S. Chaer, R. H. Bishop, and J. Ghosh, “A mixture-of-experts framework for adaptive Kalman filtering,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 27, 1997.

[11] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, “Adaptive mixtures of local experts,” *Neural Comput.*, vol. 3, no. 1, 1991.

[12] S. E. Yuksel, J. N. Wilson, and P. D. Gader, “Twenty years of mixture of experts.” *IEEE Trans. Neural Netw. Learning Syst.*, vol. 23, no. 8, pp. 1177–1193, 2012.

[13] V. Ramamurti and J. Ghosh, “On the use of localized gating in mixture of experts networks,” 1998.

[14] L. Xu, M. I. Jordan, and G. E. Hinton, “An alternative model for mixtures of experts,” in *Advances in Neural Information Processing Systems*, 1995.

[15] M. I. Jordan and L. Xu, “Convergence results for the EM Approach to Mixtures of Experts Architectures,” 1993.

[16] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. Wiley-Interscience, 2001.

[17] X. Boyen and D. Koller, “Tractable inference for complex stochastic processes,” in *In Proc. UAI*, 1998, pp. 33–42.

[18] P. Brisset, A. Drouin, M. Gorraz, P. Huard, and J. Tyler, “The paparazzi solution,” in *Micro Aerial Vehicles*, 2006.