Multisource data fusion for bandlimited signals: a Bayesian perspective[1]

A. Jalobeanu[*] and J. A. Gutiérrez[*]

[*]*LSIIT (UMR 7005 CNRS-ULP), MIV team, PASEO group @ ENSPS, Illkirch – France*

Abstract. We consider data fusion as the reconstruction of a single model from multiple data sources. The model is to be inferred from a number of blurred and noisy observations, possibly from different sensors under various conditions. It is all about recovering a compound object, signal+uncertainties, that best relates to the observations and contains all the useful information from the initial data set.

We wish to provide a flexible framework for bandlimited signal reconstruction from multiple data. In this paper, we focus on a general approach involving forward modeling (prior model, data acquisition) and Bayesian inference. The proposed method is valid for n-D objects (signals, images or volumes) with multidimensional spatial elements. For the sake of clarity, both formalism and test results will be shown in 1D for single band signals. The main originality lies in seeking an object with a prescribed bandwidth, hence our choice of a B-Spline representation. This ensures an optimal sampling in both signal and frequency spaces, and allows for a shift invariant processing.

The model resolution, the geometric distortions, the blur and the regularity of the sampling grid can be arbitrary for each sensor. The method is designed to handle realistic Gauss+Poisson noise.

We obtained promising results in reconstructing a super-resolved signal from two blurred and noisy shifted observations, using a Gaussian Markov chain as a prior. Practical applications are under development within the SpaceFusion project. For instance, in astronomical imaging, we aim at a sharp, well-sampled, noise-free and possibly super-resolved image. Virtual Observatories could benefit from such a way to combine large numbers of multispectral images from various sources. In planetary imaging or remote sensing, a 3D image formation model is needed; nevertheless, this can be addressed within the same framework.

Keywords: Model-based data fusion, uncertainties, generative models, inverse problems, signal reconstruction, super-resolution, resampling, B-Splines

1. INTRODUCTION

We can study multi-source data fusion from three different perspectives: data enhancement, decision making and optimal data reduction. The first two define the most common classification, as some seek to improve upon the data quality [16] through the simultaneous use of multiple observations, while others aim at making decisions [15] guided by multiple, more or less compatible measurements. However, we place ourselves in a different, third category, wishing to construct an object that reduces the dimensionality of the data set while conserving the maximum amount of information. This implies (but is not equivalent to) data enhancement, since the goal is to embed all measurements in a single model, minimizing the information loss, the noise contribution and the redun-

dancy. This should also help any subsequent decision making, as data analysis is greatly simplified through the use of a single object instead of multiple, heterogeneous sources.

In this study, we focus on N-dimensional signals observed through multiple instruments, and we aim at reconstructing such signals through a probabilistic framework, with sampling theory, geometry and noise modeling as basic ingredients.

To avoid losing information, it is necessary to build a multivariate probability distribution as the fused object, since the observations are actually realizations of random variables. Restricting to a signal-like, deterministic object would imply an obvious loss not only of local uncertainties, but also of correlations between interacting variables in the reconstructed signal. Indeed, initial uncertainties originating from the noise process propagate through to the very final stage of the processing, and sources are redistributed into the new object so that final variables become entangled. The uneven contributions of the observations to each final variable, as well as the possible spatial dependence of the noise, are strong evidence that model uncertainties are useful information that ought to be retained during the fusion.

2. A BAYESIAN APPROACH TO DATA FUSION

2.1. Modeling uncertainty: Evidence vs. Bayesian theory

Uncertainty, as distinguished from imprecision, arises as a result of random process modeling or when there is missing data, where the lack of information induces some belief. What formalism is appropriate to treat evidence in the image fusion domain? Two approaches can be considered right now: Bayesian theory [5] and Belief Function theory [12] (a.k.a. Dempster-Shafer theory or theory of Evidence [6]).

The primary advantage of the Bayesian paradigm is that it is well developed, and evidence is treated as a probability density function (pdf). By applying Bayes' rule, we know how to compute a posteriori knowledge based on the a priori knowledge and on the conditional probability of the evidence.

As a fusion process combines the observations from different sources into a single coherent perception of a scene, then the reduction of the uncertainties is expected; hence observations could be treated as evidence and evidential reasoning could be used to infer the original scene. The Evidence theory exhibits two main advantages: it explicitly represents ignorance, and does not require a priori knowledge. Indeed, evidence is represented as a belief function allowing any portion of the belief mass to be explicitly assigned to ignorance; when evidence is acquired, belief replaces ignorance.

Nevertheless, no modeling of ignorance is explicitly needed within our data fusion problem. Moreover, a priori knowledge is needed. In fact, the nature of our fusion process, as a recursive method, makes us take advantage of prior information at each update step. Besides, we always try to stay away from computationally intractable issues while inferring evidence within the Bayesian framework. This situation could worsen while distributing belief to the power set of hypothesis considered within the Belief Function framework. The straightforward choice is then the Bayesian theory.

2.2. A generative model-based approach

From the Bayesian point of view [5], all variables (including parameters) are random variables. We consider here the general case of real-valued variables. They are either unknown or observed. Among all unknown variables, some of them are quantities of interest (here, the single fused object we want to determine), others are nuisance variables and they need to be integrated out. The likelihood term describes how the observed data are formed given all other quantities. To fully specify the joint probability density function (pdf) for all variables, priors are required. We could then summarize the Bayesian treatment as the definition of a likelihood through the description of the experimental setup, the specification of priors given all available knowledge, and the marginalization [5, 2] to eliminate all unwanted variables.

Thus, the data fusion problem can be stated as the computation of the posterior pdf of the unknown single object given all observations. If X denotes the object to be reconstructed, and $\{Y\}$ the set of observations, Bayes' rule enables us to express this posterior as $P(X \,|\, \{Y\}) = P(\{Y\} \,|\, X)P(X)/P(\{Y\})$ where $P(\{Y\} \,|\, X)$ is the likelihood, $P(X)$ the prior, and the last term is referred to as evidence and does not depend on X. Therefore the joint pdf and the posterior are proportional. Notice that all unwanted variables are implicitly integrated out.

The model variables involved in this approach are usually linked to each other through deterministic relations or, stochastically, through conditional probability densities. Thus, building the likelihood function is not always an easy task. However, knowing the hierarchy between variables (e.g. the sequence of physical processes leading from initial parameter values to the final data set) helps constrain this function, and can be nicely encoded in a graph structure, hence our resort to graphical models. Directed graphical models or Bayesian networks [8] enable us to fully define a joint pdf for even complex models: the joint is factorized into a product of conditional pdfs, represented by all nodes with a set of converging arrows, and prior pdfs related to the remaining nodes.

We use a generative model to build the likelihood, rather than choosing popular or ad-hoc functions; it is crucial to accurately understand how the data were generated (through modeling each type of instrument used to produce each observation) in order to effectively invert the observation process and reconstruct what was actually the origin of the recorded set of measures. In this framework, we only try to introduce knowledge we are certain of, for instance the experimental setup and some well-calibrated parameter values. Precise statistical properties can be encoded in graph structures, such as the conditional independence of observations given the model, or peculiar dependence patterns within the model itself. Specifying the structure of the graph only specifies the dependence between variables. Choosing the shape of the density functions strongly constrains the problem and can sometimes bring more knowledge than intended, thus priors have to be specified with great care. Whereas the functions involved in the likelihood are derived from physics and are often well-known, the priors should be designed to only enforce objective constraints, easy to check experimentally (such as image smoothness properties, or positivity) without being too informative regarding all that is unknown. To avoid dealing with such issues, we will try to minimize the weight of the priors by minimizing the number of parameters in the model to reconstruct, to ensure that there will (almost) always be enough data points to determine each unknown variable accurately enough.

2.2.1. Bayesian networks for multiple observations

The formalism presented here is quite general and does not only apply to N-dimensional signals. Let us call observation any group of recordings of the same physical phenomenon taken through a sensor, characterized by a set of parameters related to the experimental setup (e.g. in imaging, the camera parameters including pose, blur, sensor geometry...). An observed signal Y^n is generated given an object model X, its resolution (or density) ϵ that acts as a spatial scaling parameter, and of course the observation parameters Θ^n for this particular sensor indexed by n. Now we can also constrain the image model itself by conditioning it upon a set of parameters ω. Hence the graphical model of Fig. 1. Following the related formalism, the joint pdf is then given by the product:

$$P(X, \{Y^n\}, \{\Theta^n\}, \omega, \epsilon) = P(\epsilon)P(\omega)P(X|\omega)\prod_n P(\Theta^n)P(Y^n|X, \theta^n, \epsilon) \quad (1)$$

As the goal is to get the posterior marginal, integrating with respect to Θ, ω and ϵ is required (this is also called marginalization):

$$P(X|\{Y^n\}) \propto \int P(\epsilon)P(\omega)P(X|\omega)\prod_n P(\Theta^n)P(Y^n|X, \Theta^n, \epsilon)\, d\Theta^n\, d\omega\, d\epsilon \quad (2)$$
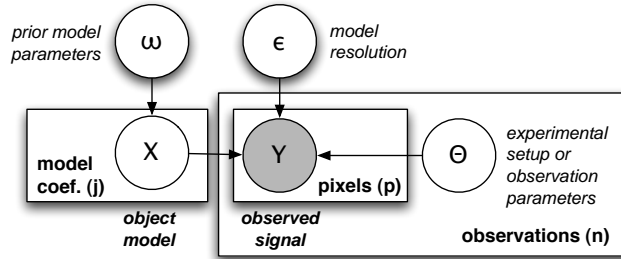

FIGURE 1. The proposed directed graphical model explaining the formation of multiple observations Y^n from a single model X (governed by ω), given experimental setups Θ^n and a model density ϵ. Nodes represent random variables (conditional pdf for converging arrows, prior pdf without incoming arrows).

2.2.2. A band-limited signal resampling scheme

We now need to precisely describe the generative model in terms of signal transformation, from an input X to each observation Y. We are making an important assumption here: the signals we consider are band-limited since they are recorded using instruments with limited bandwidth (e.g. finite-size optical systems in imaging) [7]. Indeed, there is no point in trying to recover signals with arbitrary spectral content when performing data fusion, as the goal is to combine observations, not to deconvolve or analyze them. We explicitly use the point spread function (PSF) denoted by h, related to the transfer function of the system through a Fourier transform. Assuming a ground truth or underlying scene T, the formation of the observed signal mean I is modeled as follows:

- Geometric mapping from model space to sensor space (both included in $\mathbb{R}^N$) denoted by f, the respective spatial locations are x and u. It is defined by the observation parameters θ, α and the model density ϵ. We use the inverse $g = f^{-1}$ so as to express the mapped T as $T \circ g(u) = T(g(u))$ in the sensor space.

- Convolution with the possibly spatially variable PSF h: we have

$$(T \circ g) \star h(u) = \int_{\mathbb{R}^N} T(g(v)) \, h_u(u-v) \, dv \tag{3}$$

- Sampling using a pixel grid denoted by π: point π_p for each detector (pixel) p.

$$I_p = (T \circ g) \star h(\pi_p) \tag{4}$$

Let us start with the properties that the fusion method must have in order to behave as expected in well-known special cases. The goal is to constrain the proposed signal resampling scheme. Though a formal proof is beyond the scope of this paper, we provide an explanation based on induction to help understand the generative model design. 1) We want to resort to a simple weighted average without deblurring when inputs are ideally sampled, aligned and obtained using the same sensor and instrument geometry; 2) We need our fusion framework to be a generalization of state of the art interpolation in the presence of noise [9, 14] when inputs are not aligned; 3) We wish to handle non-uniform spaced samples in the same way [1]. 4) Arbitrary geometries shall be allowed while conserving the energy (i.e. photometry conservation for images).

Property 1. It arises from the band-limiting requirement, stating that the end result can not be arbitrarily oversampled, and also from the information conservation condition related to sampling theory, i.e. the reconstructed signal shall not be undersampled [7]. Some authors suggest that B-Splines provide a near-optimal sampling [13] since their spatial footprint is finite and small, while having a reasonably limited bandwidth (though not finite). An obvious advantage of B-Splines over other sampling kernels is their fast and accurate implementation when it comes to interpolation [13]. These characteristics mainly motivated our choice; we retain the B-Spline of degree 3, denoted by φ, as a good compromise between accuracy and computational complexity. Then our target is actually the continuous, (almost) band-limited signal $F = T \star \varphi$; we aim at reconstructing object Spline coefficients denoted by L such that:

$$F(x) = \sum_j L_j \varphi(x-j) \quad \text{with} \quad j \in \mathbb{Z}^N \tag{5}$$

where j refers to the model space discretization. We will rather aim at the sampled version of F, denoted by X such that $X_p = F(p)$, containing exactly the same information since F it is nearly band-limited by definition. It is expressed through a discrete convolution $X = L \star s$ where $s_j = \varphi(j)$. Thus we can easily use popular prior models defined for discrete signals, governed by some parameters ω; for instance the positivity of X (due to physical reasons) can be enforced whereas the coefficients L can take negative values as required by interpolation theory. Hence the left part of the graph on Fig. 2.

In the special case of property 1, the only way to get a linear combination of observations at each point is if $X = I$ (no intensity scaling so all the I^n are a copy of X). Therefore each I_p is a linear combination of L_j by using the definition of X: we have $I_p = \sum_j \lambda_{pj} L_j$, and in this special case $\lambda_{pj} = h(p-j)$ where $h = \varphi$.

The generalization is straightforward: let I be the blurred version of T such that $I_p = T \star h(p)$. If h is band-limited such that it can be written as $h(u) = \sum_j c_j \varphi(u-j)$, then by expanding h we get $I_p = \sum_j h(p-j)L_j$. The band-limiting of h in the model space is essential. We show in the following paragraphs how to enforce this constraint.

Property 2. If an observation is shifted by b, we want to resort to Spline interpolation to cancel the shift. We set $X_p = \mathcal{S}[I](p+b)$ where $\mathcal{S}[I]$ stands for the Spline interpolation of I. Since shifting through interpolation is invertible, we can write $I_p = \mathcal{S}[X](p-b) = F(p-b)$ which gives $\lambda_{pj} = h(p-j-b)$.

Property 3. We wish to allow an arbitrary sensor space discretization, using for each sample real coordinates π_p instead of integers p. This amounts to reconstructing a signal from nonuniform samples by minimizing a sum of error terms for each π_p and a smoothness term as in [1] when there is a single input signal.

Property 4. In the special case above we have $g(u) = u - b$, but one could easily imagine more complex functions so as to include scaling and rotations (i.e. $Au + b$), and also nonlinear transforms such as distortions or perspective effects. Let us denote by W_θ and b_θ the linear transform from local to world frame, and by g_α^s the transform from sensor to local frame, respectively parametrized by so-called external and internal observation parameters θ and α. The model density ϵ acts as a global scaling factor.

$$g(u) = \epsilon \left(W_\theta g_\alpha^s(u) + b_\theta \right) \tag{6}$$

We assume g has a linear behavior over the footprint of h. Let J_u be the Jacobian matrix of g at location u so that $g(u+w) \simeq g(u) + J_u w$. We express the convolution (3) in the model space through a change of variables, and use the linear approximation to get

$$\int T\big(g(v)\big) h_u(u-v)\, dv \simeq \left| J_u^{-1} \right| \int T(v') h_u\Big(J_u^{-1}\big(g(u) - v'\big) \Big) dv' \tag{7}$$

The Jacobian determinant ensures the conservation of energy. Now we need to expand the warped PSF h using B-Splines as seen previously so we can revert to Spline coefficients L. The PSF bandwidth must fulfill the sampling conditions (sample spacing = 1) even after remapping through g, which constrains the choice of the model density ϵ to be higher than the Nyquist rate. Thus, fusing undersampled signals requires achieving super-resolution by increasing the model density accordingly. Some calculus gives

$$(T \circ g) \star h\,(u) = \sum_j L_j \left| J_u^{-1} \right| h_u\Big(J_u^{-1}\big(g(u) - j\big) \Big) \tag{8}$$

Let us sample at $u = \pi_p$ and adopt a simpler notation $\Delta_p^n = J_{\pi_p}^{-1}$ and $h_p^n = h_{\pi_p}$ for each observation indexed by n; each n-th mean observation at pixel p now writes

$$I_p^n = \sum_j \lambda_{pj}^n L_j \quad \text{where} \quad \lambda_{pj}^n = \left| \Delta_p^n \right| h_p^n\Big(\Delta_p^n\big(g(\pi_p) - j\big) \Big) \tag{9}$$

2.2.3. The proposed generative model

The stochastic nature of the recorded signal comes from the observation noise. Its mean is I_p^n. Its pdf can be modeled by a succession of Poisson counting, Gaussian readout and uniform quantization processes, whose pdfs are convolved, and reasonably approximated by a Gaussian additive process whose variance is a linear function of the mean. We assume that given the mean, all observations are independent, between sensors n as well as elementary detectors or pixels p. Therefore we have:

$$P(\{Y^n\} \,|\, \{I^n, \tau^n, \sigma^n\}) = \prod_n \prod_p \frac{1}{\sqrt{2\pi}\, v_p^n} e^{-\left(Y_p^n - I_p^n\right)^2 \big/ 2 v_p^n} \quad \text{with } v_p^n = \tau^n I_p^n + (\sigma^n)^2 \tag{10}$$

where I^n is the n-th observed signal as in the rendering equation (9), deterministic function of L and, through λ^n, of (π^n, h^n) and, through g, of $(\alpha^n, \theta^n, \epsilon)$. In the full graphical model of Fig. 2, all deterministic relations of Eqns. (6) and (9) are expressed through Dirac distributions, such as $\delta(I^n_p - \sum_j \lambda^n_{pj} L_j)$ for instance, the object X being linked to L through $\delta(X - L \star s)$. This obviously simplifies the integration in Eqn. (2). Now the entire observation model is defined and parametrized by $\Theta^n = (\theta, \alpha, h, \sigma, \tau)^n$.

In order to express the full joint pdf (1) we need to specify $P(X \,|\, \omega)$, expressing the prior object model. For the sake of simplicity, we consider a first order Gaussian Markov Random Field (Markov chain if $N{=}1$) parametrized by a single smoothness parameter ω. It could be replaced by more application-specific priors (e.g. point sources over a smooth background for astronomical images, or wavelet-based models in remote sensing). Let Z_ω denote a normalizing constant, and $\{q \sim p\}$ the first order neighbors of p:

$$P(X \,|\, \omega) = \frac{1}{Z_\omega} e^{-\Phi_\omega(X)} \quad \text{where} \quad \Phi_\omega(X) = \omega \frac{1}{2^N} \sum_p \sum_{\{q \sim p\}} (X_p - X_q)^2 \qquad (11)$$

If the PSF, the internal parameters α, the sampling grid and the noise variance are well-calibrated, and ϵ fixed above Nyquist rate, then the joint posterior for all unknowns is

$$P(X, \omega, \{\theta^n\} \,|\, \{Y^n\}) \propto P(\omega) \frac{1}{Z_\omega} e^{-\Phi_\omega(X)} \prod_n P(\theta^n) \prod_p \frac{1}{v^n_p} e^{-\left(Y^n_p - I^n_p\right)^2 / 2v^n_p} \qquad (12)$$
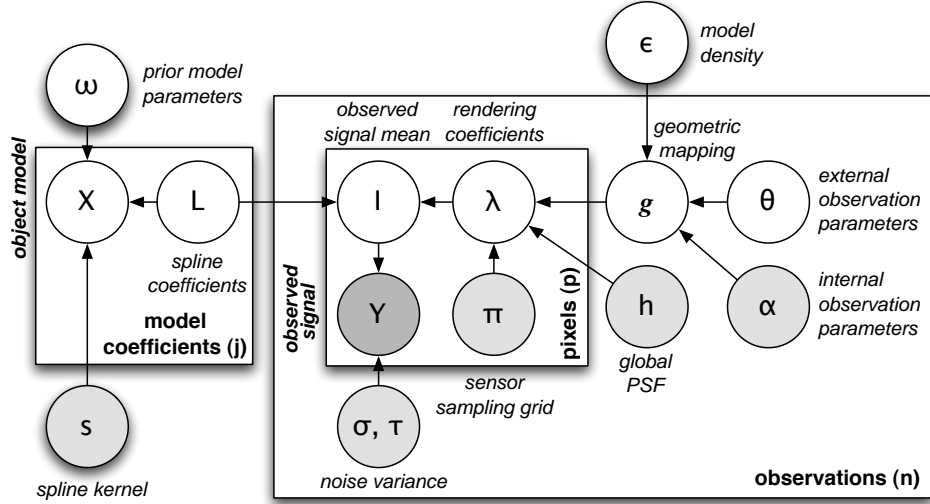

FIGURE 2. Expanded graphical model of Fig. 1 displaying the generative model for observations Y, the rendering of I and the B-Spline coefficients L. Shaded nodes represent variables having a fixed value.

3. DATA FUSION AS AN INVERSE PROBLEM

We address multisource data fusion as an inverse problem, the direct problem being the observed signal formation described by the generative model of Section 2.2. If enough data points per unknown variable are available, the problem is well-posed. However this can never be guaranteed, unless we allow some loss of information by decreasing the model density ϵ. In order to maintain a minimal sampling rate everywhere, some areas may be under-determined if the mapping g is not linear, hence the use of a prior model.

3.1. The supervised inference scheme

Within a supervised framework, both prior and observation parameters are assumed calibrated. Then we aim at computing the conditional pdf $P(X \mid \{Y^n, \theta^n\}, \omega)$. We approximate it by a Gaussian distribution; let us first focus on its mean or mode, obtained by finding $\hat{X}$ that maximizes this pdf. To simplify, we rather work with L, and get the end result as $\hat{L} \star s$. It all comes down to minimizing an energy term U:

$$U(L) = -\log P(L \mid \{Y^n, \theta^n\}, \omega) = \Phi_\omega(L \star s) + \sum_{n,p} \frac{\left(Y_p^n - I_p^n\right)^2}{2v_p^n} + \log v_p^n + \text{const.} \quad (13)$$

which can be done iteratively by a quasi-Newton method, consisting of optimizing a quadratic approximation of U through a conjugate gradient search [11] at each step. Indeed, at each step we fix v_p^n in (10) by setting I_p^n using the current estimate of L. This is not needed for a pure Gaussian noise (τ=0) in which case the whole problem is linear.

The initialization is performed using the rendering coefficients λ as local weights:

$$L_j^0 = \left(\sum_{n,p} \lambda_{pj}^n Y_{pj}^n\right) \bigg/ \left(\sum_{n,p} \lambda_{pj}^n\right) \quad (14)$$

If we use a vector notation (the components being the model coefficients indexed by j) such that $\Phi_\omega = \omega X^t D^t D X$ with $X = SL$, then the gradient and the Hessian write:

$$\nabla_L U = 2\omega \, (DS)^T (DS) \, L + \sum_{n,p} \frac{1}{v_p^n} \left(\lambda_p^n \cdot L - Y_p^n\right) \lambda_p^n \quad (15)$$

$$\nabla_L^2 U = 2\omega \, (DS)^T (DS) + \sum_{n,p} \frac{1}{v_p^n} \left(\lambda_p^n\right) \left(\lambda_p^n\right)^T \quad (16)$$

3.2. Computing, approximating and propagating uncertainties

Within the Gaussian approximation, uncertainties on the estimate $\hat{X}$ are represented by the covariance matrix Σ_X, whose inverse is the Hessian $\nabla_X^2 U$. From $X = SL$ we get

$$\Sigma_X^{-1} = \left(S^{-1}\right)^T \left(\nabla_L^2 U\right) \left(S^{-1}\right) = 2\omega \, D^T D + \sum_{n,p} \frac{1}{v_p^n} \left(S^{-1}\lambda_p^n\right) \left(S^{-1}\lambda_p^n\right)^T \quad (17)$$

Applying the inverse S^{-1} takes great advantage of the B-Spline theory, which provides a computationally effective algorithm [13]. The resulting matrix can be very large, but it is sparse, due to the small footprint of h compared to the size of the signal.

Due to the slow decay of the PSF in some cases, the coefficients λ defined in (9) can induce many small but nonzero entries in the matrix Σ_X^{-1}; for practical reasons, we would like to perform a so-called covariance simplification, in order to make it arbitrarily sparse. Indeed, providing uncertainties in addition to the fusion result ought not to overwhelm the end user by fairly large numbers of extra parameters. Moreover, propagating them through to subsequent processing steps (such as updating, or data analysis) could result in an uncontrolled growth of model complexity. One can possibly minimize a distance between the posterior distribution and a simplified distribution, subject to constraints – for instance, the simplified inverse covariance $\tilde{\Sigma}_X^{-1}$ has the same Markov structure as the prior $D^t D$, i.e. only near-diagonal elements are nonzero. Optimization methods for sparse matrices are still subject to intensive research.

The recursive nature of our fusion methodology allows us to perform sequence updates as long as new observations are available. Bayesian updating allows us to use the current (approximate) posterior pdf as a prior when making inference from new observations. Thus, data can be processed in groups, to reduce memory requirements when dealing with large data sets, or to update an existing fusion result without having to restart the entire inference from the beginning. The new prior writes:

$$P^*(X) \propto e^{-\Phi^*(X)/2} \quad \text{where} \quad \Phi^*(X) = X^T \tilde{\Sigma}_X^{-1} X \tag{18}$$

3.3. Towards a fully automated procedure

There are several ways to make data fusion unsupervised (when neither ω nor θ^n are known), all starting with the marginalization of Eqn. (2) but applying different approximations. When the dimensionality of the parameters is small compared to the number of variables in X, their posterior $P(\omega, \{\theta^n\} \,|\, Y)$ is very peaked around the optimum $(\hat{\omega}, \{\hat{\theta}^n\})$, which leads to the so-called empirical Bayesian formalism [5, 9]:

$$P(X \,|\, \{Y^n\}) = \int P(X \,|\, \{Y^n, \theta^n\}, \omega) P(\omega, \{\theta^n\} \,|\, Y) \, d\omega \, d\{\theta^n\} \simeq P(X \,|\, \{Y^n, \hat{\theta}^n\}, \hat{\omega}) \tag{19}$$

Finding this optimum requires integrating (12) with respect to X (or L, equivalently):

$$\left(\hat{\omega}, \{\hat{\theta}^n\}\right) = \arg \min_{\omega, \{\theta^n\}} \left(\frac{Z_{\omega\theta}}{Z_\omega} P(\omega) \prod_n P(\theta^n) \right) \text{ where } Z_{\omega\theta} \equiv \int e^{-U(L)} \, dL \tag{20}$$

A Laplace approximation [10] could be used, the optimum and Hessian being provided by the supervised inference algorithm of section 3.1. The main difficulty lies in computing the Hessian determinant, whereas Z_ω can be calculated for the model of Eqn. (11) [9]. Neglecting the contribution of this determinant provides a rather simple, nested algorithm that runs a supervised inference within each step of the parameter search, giving in the end not only the parameters, but also the fusion result with uncertainties. The parameter search is nonlinear; it can be done through gradient methods [11], or alternatively by Expectation-Maximization [5], simpler but more sensitive to local minima.

4. POSSIBLE APPLICATIONS AND EXTENSIONS

The proposed formalism can be applied to 1D problems (as demonstrated by our preliminary results). For instance spectrum fusion in physics would combine spectra measured with different instruments and thus improve accuracy and faint source detection.

In 2D, the new fusion method aims at bringing new solutions to image mosaicing, co-addition, and super-resolution with an unprecedented management of uncertainties and an explicit handling of geometry, blur and noise statistics. Application areas include astronomy [4] where huge amounts of data are available within virtual observatories, as well as remote sensing and planetary imaging [3]. A common issue, particularly well-suited to the new fusion framework, is the optimal combination of images taken at different resolutions, in various viewing conditions (and not perfectly registered), with specific noise properties and possibly missing data. The extension to multispectral data is in progress; it should provide a unified framework for not only spatial resampling as shown in this paper, but also spectral resampling via a fully Bayesian approach.

5. FIRST RESULTS: SUPER-RESOLUTION FOR SIGNALS

We made a 1D experiment, using an analytic function as ground truth (two peaks, with an additive sinusoidal oscillation) as shown on Fig. 3. Two observations (left) were generated via a linear geometric mapping, a Gaussian blur and a regular sampling grid such that the sampling rate (with respect to the blur) was 0.6 times the Nyquist rate, producing obvious aliasing effects. Therefore we aimed at a super-resolved output to recover the aliased frequencies (factor 2 or $\epsilon = 0.5$, inducing a slight deblurring). The supervised method was used. Fusion allowed the recovery of the oscillations without too much noise amplification, thanks to the Markovian prior model. Uncertainties were computed: the inverse covariance matrix was inverted numerically while the error bars (uncertainty envelope) correspond to 95% confidence and are given by diagonal elements.
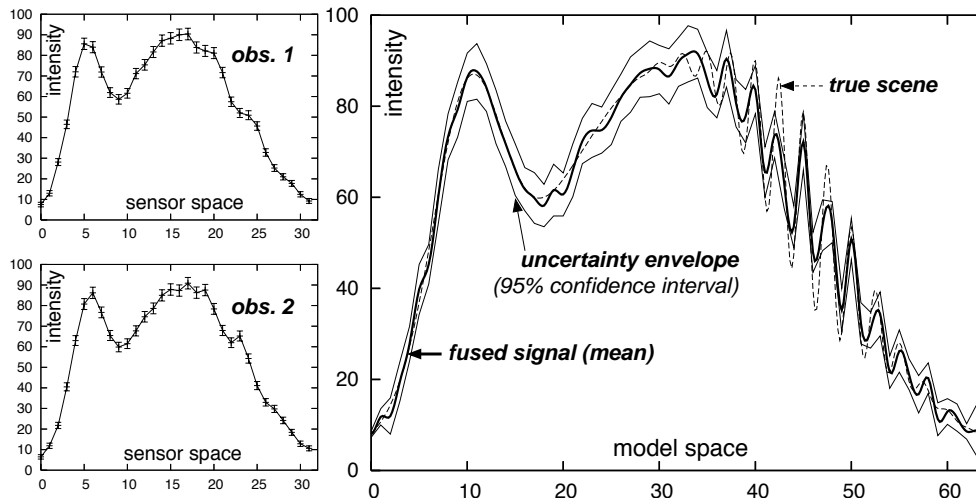
FIGURE 3. Super-resolved fusion of two undersampled 1D signals (shifted, blurred and noisy).

REFERENCES

1. M. Arigovindan et al. Variational image reconstruction from arbitrarily spaced samples: A fast multiresolution spline solution. *IEEE Trans. on Image Processing*, 14(4), 2005.
2. J.O. Berger, B. Liseo, and R. Wolpert. Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science*, 14(1), 1999.
3. P. Cheeseman et al. Super-resolved surface reconstruction from multiple images. In G. R. Heidbreder, editor, *Maximum Entropy and Bayesian Methods*, 1996.
4. A.S. Fruchter and R.N. Hook. Drizzle: A method for the linear reconstruction of undersampled images. *Publications of the Astronomical Society of the Pacific (PASP)*, 114(792), 2001.
5. A. Gelman, J.B Carlin, H.S Stern, and D.B Rubin. *Bayesian Data Analysis*. Chapman & Hall, 1995.
6. D.L. Hall and S.A. McMullen. *Mathematical Techniques in Multisensor Data Fusion*. Artech, 2004.
7. A.K. Jain. *Fundamentals of digital image processing*. Prentice Hall, 1989.
8. M. I. Jordan, editor. *Learning in graphical models*. MIT Press, 1998.
9. D.J.C. MacKay. Bayesian interpolation. *Neural Computation*, 4(3), 1992.
10. D.J.C. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge, 2003.
11. W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 2nd edition, 1993.
12. P. Smets. Belief functions on real numbers. *International Journal of Approximate Reasoning*, 2005.
13. P. Thévenaz et al. Interpolation revisited. *IEEE Trans. on Med. Imaging*, 19(7), 2000.
14. M. Unser and T. Blu. Generalized smoothing splines and the optimal discretization of the Wiener filter. *IEEE Trans. on Signal Processing*, 53(6), 2005.
15. P. K. Varshney. *Distributed Detection and Data Fusion*. Springer, 1996.
16. L. Wald. Some terms of reference in data fusion. *IEEE Trans. on Geosci. Remote Sens.*, 37(3), 1999.