A BAYESIAN MULTI-FRAME IMAGE SUPER-RESOLUTION ALGORITHM USING THE GAUSSIAN INFORMATION FILTER

Matthew Woods and Aggelos Katsaggelos

Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL USA

ABSTRACT

Multi-frame image super-resolution (SR) is an image processing technology applicable to any digital, pixilated camera that is limited, by construction, to a certain number of pixels. The objective of SR is to utilize signal processing to overcome the physical limitation and emulate the "capabilities" of a camera with a higher-density pixel array. SR is well known to be an *ill-posed* problem and, consequently, state-of-the-art solutions approach it statistically, typically making use of Bayesian inference. Unfortunately, direct marginalization of the posterior distribution resulting from the Bayesian modeling is not analytically tractable. An approximation method, such as Variational Bayesian Inference (VBI), is a powerful tool that retains the advantages of statistical modeling. However, its derivation is tedious and model specific. In this paper, we propose an alternative approximate inference methodology, based upon the well-established, Gaussian Information Filter, which offers a much simpler mathematical derivation while retaining the statistical advantages of VBI.

Index Terms— Super-Resolution, Image-Processing, Inverse Problems, Remote Sensing, Photogrammetry

1. INTRODUCTION

Multi-frame image super-resolution (SR) is an image processing technology applicable to any digital, pixilated camera that is physically limited, by construction, to sample a scene with a discrete, m x n pixel array. The straightforward objective of SR is to utilize signal processing to overcome this physical limitation of the m x n array and emulate the "capabilities" of a camera with a higher-density, $Mm \times Mn \quad (M > 1)$ pixel array. M is the "magnification" factor" of the SR algorithm. The exact meaning of "capabilities", in the preceding sentence, is application dependent. As the name implies, multi-frame SR techniques take multiple low-resolution (LR) frames as input and combine them in some way to produce a high-resolution (HR) output. In doing so, SR algorithms capitalize on the fact that high spatial-frequency content in the scene is present in the LR images, but degraded due to aliasing resulting from limited sampling density.

SR is well known to be an *ill-posed* problem. Consequently, state-of-the-art solutions treat SR as a matter of *inferring* the HR image based not only on the LR input images but also on prior information about the HR image, frame-to-frame motion, etc. A hierarchical Bayesian framework provides a powerful and flexible means to incorporate all of this information into the SR problem [1,2].

Unfortunately, direct integration of the *posterior* distribution, resulting from the Bayesian modeling, in order to get the expected value of the HR image, is not analytically tractable. This leads to three potential solutions, namely, i) Maximum a Posteriori (MAP), ii) Sampling methods, and iii) Approximate inference.

Due to its simplicity, the MAP method is the most commonly seen in practice. A typical example is the algorithm of Farsiu, et. al [3]. However, MAP solutions can't exploit the full potential offered by probabilistic modeling as only the posterior mode is sought [4]. As the forward model becomes more complex, either due to more complex image priors or simultaneous estimation of other parameters such as camera motion, finding the mode can become very sensitive to local minima.

Sampling methods, such as Markov Chain Monte Carlo (MCMC) or Hamiltonian Monte-Carlo (HMC) [4], work in principle but are computationally expensive. They do, however, lend themselves to mass parallelization. So, it is possible that with the increase in computing power, particularly using devices such as GPUs that exploit mass parallelization, these methods may see more frequent use. However, we do not consider them in this paper.

Approximate inference methods, such as Variational Bayesian Inference (VBI) which is well described in [4], allow us to retain the advantages of statistical modeling and obtain the full posterior in a computationally efficient manner. The VBI approach is used in several state-of-the-art solutions [1,2,6]. The VBI methods are powerful; however, their derivation is advanced, tedious, and model specific [7,8]. It is difficult to quickly experiment with different models as the VBI solution must be carefully re-derived each time. This difficulty limits the general appeal of the, statistically superior, VBI solution and frequently leads practitioners to abandon it in favor of conceptually simpler, yet less powerful methods, methods such as the single point MAP solution. In this paper, we propose an alternative approximate inference methodology based upon the well-established, Gaussian Information Filter (IF). The IF is a dual-form of the popular Kalman filter and offers a much simpler mathematical derivation while retaining the statistical advantages of VBI. We believe this alternative will make Bayesian inference solutions more accessible to the general practitioner.

The remainder of the paper is organized as follows. In section 2, we introduce the hierarchical Bayesian formulation of the SR problem. In section 3, we derive our solution using a combination of the IF and the expectation-maximization (EM) technique to create an "IFEM" algorithm. In section 4, we compare the IFEM to the state-of-the-art using image data captured from a drone. In section 5, we conclude the paper.

2. HIERARCHIAL BAYESIAN MODEL

In the following, we use the formulation presented in [1,2]. This model assumes that the imaging process has captured L LR images y_k from an unknown HR image x. Note that, in this formulation, both the LR and HR images are already in the discrete, pixelated domain. The LR images y_k and the HR image x consist of a total of N and M^2N pixels, respectively, where the value M > 1 is the magnification factor representing the increase in resolution. In order to represent the problem compactly in matrix-vector notation, the images y_k and x are arranged in lexicographical order as N x 1 and MN x 1 vectors, respectively. The imaging process model includes warping, blurring (MTF), noise, and down-sampling as

$$y_k = AH_k C(s_k)x + n_k, \tag{1}$$

where A is the $N \times M^2 N$ downsampling matrix, H_k is the $M^2 N \times M^2 N$ blurring matrix, $C(s_k)$ is the $M^2 N \times M^2 N$ warping matrix generated by the image motion vector s_k , and n_k is the $N \times 1$ acquisition noise.

Given (1), the SR problem is to find the best estimate of the HR image x from the set of LR images y_k using prior knowledge about $C(s_k)$, n_k , and x. Note, at this point, we have not had to specify the specific functional form of the warping model $C(s_k)$ or the dimensionality of the set of warping parameters $\{s_k\}$. We can now define the joint posterior of all unknowns as

$$Pr[x, \{s_k\}, \{\beta_k\}, \alpha | \{y_k\}, \{\Omega_{sk}\}]$$

$$\propto \prod_{k=1}^{L} \left(\beta_k^{\frac{N}{2}} exp\left(-\frac{\beta_k}{2} \| y_k - (2)\right)\right)$$

$$AH_kC(s_k)x\|^2 \exp\left(-\frac{1}{2}s_k^T\Omega_{sk}s_k\right)Pr[\beta_k]\right)Pr[x|\alpha]Pr[\alpha],$$

where Ω_{sk} is the measurement precision matrix of the warping parameters for LR frame k, β_k is a hyper-parameter

for the likelihood of measured image k, α is a hyperparameter for high-resolution image prior model, $Pr[x|\alpha]$ represents the HR image prior model, and the terms $Pr[\beta_k]$ and $Pr[\alpha]$ represent hyper-priors on the hyperparameters. A common image prior model, $Pr[x|\alpha]$, is the Total Variation (TV) prior, which is used for image reconstruction problems due to its inherent ability to retain sharp gradients at image edges [2], given by

$$Pr[x|\alpha] =$$

$$^{M^2N/2}exp\left(-\frac{\alpha}{2}\sum_{i=1}^{M^2N}\sqrt{\Delta h_i^2 + \Delta v_i^2}\right),$$
(3)

where Δh_i and Δv_i are the horizontal and vertical gradients, respectively, for pixel *i* in the HR image *x*. Δh_i and Δv_i may be found using any reasonable gradient estimator. Another common image prior is the Gaussian, Simultaneous Auto-Regressive (SAR). However, it is known to not preserve image edges as well as the TV prior.

α

The hyper-priors, $Pr[\beta_k]$ and $Pr[\alpha]$, are commonly modeled using either the uninformative distribution, in which case they have to be complete estimated from the data itself, or as Gamma distributions [2]. The Gamma distribution is convenient from an analytic tractability perspective in that it is conjugate to the normal distribution [5]. A conjugate prior distribution is one that results in a posterior distribution having the same functional form as the prior. The ability of the algorithm to automatically learn the hyper-parameters from the data, either with an uninformative prior or some guidance via a Gamma prior, is a powerful capability. Other popular, non-Bayesian, SR methods leave the hyperparameter estimation to the user which requires a long parameter-tuning process and can limit the applicability of the solution, as described in [2].

3. DERIVATION OF THE PROPOSED INFORMATION-FILTER / EXPECTATION-MAXIMIZATION (IFEM) SOLUTION

Prior work on VBI for SR has shown that, without prior assumption, the distribution models which minimize the Kullback-Leibler (KL) divergence for both the unknown high-resolution image x as well as unknown image registration parameters $\{s_k\}$ are, indeed, Gaussian distributions [1,2]. Given this result, we propose to solve the SR problem using the well-established tools and theory surrounding Bayesian Gaussian filters [9]. The extended Kaman filter (EKF), which supports non-linear models via local linearization, is the most popular of these and has been previously examined for SR [10]. However, the EKF solution creates several issues. We offer two changes to address these issues. The most significant issue arises from the required size of the state covariance matrix. An HR image with M^2N pixels requires a very large covariance matrix with (M^4N^2) elements. However, even though the covariance matrix is large, it is very sparse. The actual required size can be judged by estimating the number of likely non-zero correlations for each image pixel. Each pixel will be correlated to its immediate neighbors due to image motion and blur. It will also be correlated to each element in the set of warping parameters $\{s_k\}$. If there are an average of Q of these non-trivial correlations per pixel, we expect the total number of non-zero elements of the covariance matrix to be $QM^2N \ll (M^4N^2)$. Consequently, by using modern, numerical linear algebra software, which is able to utilize sparse matrices, we are able to overcome this limitation. Support for sparse matrix operations is available in such packages as Matlab and NVIDIA's Cuda library [11].

Secondly, the EKF, in its native form, has numerical difficulties with the SR problem. The EKF is characterized by a relatively straightforward time update step and a more difficult measurement update step. In contrast, its dual formulation, the information filter (IF) has a difficult time update step and a relatively straightforward measurement update step [9]. In the SR formulation, we are effectively only using the measurement update of the Bayesian filter. Therefore, the IF turns out to be a computationally simpler and more numerically stable solution. We also investigated the Unscented Kalman Filter (UKF) which is a Gaussian filter closely related to the EKF [9,12] and has some precedent in image enhancement applications such as film-grain removal [13]. In principle, the UKF is better able to handle non-linear models than the EKF. There is, however, no existing dual formulation of the UKF akin to the IF and we find that the UKF suffers from the same numerical difficulties with SR as the EKF.

3.1. Derivation of the IF Solution

We start with a general description of the IF. Given an unknown state vector θ and a series of *P* measurement residuals $\{z'_p\}$, it solves for the posterior probability distribution of the form

$$Pr[\Delta\theta|\{z'_p\}] = \frac{\prod_{p=1}^{P} Pr[z'_p|\Delta\theta] Pr[\Delta\theta]}{Pr[\{z'_p\}]},$$
(4)

where a "measurement residual" is the difference between the actual measurement and a prediction of the measurement based upon the current best-estimate of the state θ and $\Delta\theta$ is a correction to the current best-estimate of the state θ . Using $\Delta\theta$ as opposed to θ in (4) allows us to handle non-linear models through local linearization. We model the relationship between the measurement residual z'_p , the actual measurement z_p , and the state θ through

$$z'_{p} = z_{p} - h_{p}(\theta) \cong \left[\frac{\delta h_{p}}{\delta \theta}\right] \Delta \theta, \text{ and}$$

$$Pr[z'_{p}|\Delta \theta] \sim N\left(\left[\frac{\delta h_{p}}{\delta \theta}\right] \Delta \theta, Q_{p}\right),$$
(5)

where Q_p is the covariance of the measurement residual z'_p and $h_p(\theta)$ is a, in general non-linear, function that predicts the measurement, z_p , based on the current state estimate. We linearize the measurement function locally about θ by computing the Jacobian $\left[\frac{\delta h_p}{\delta \theta}\right]$.

A general, analytic solution to (4) is available for the special case that all of the above probability functions are Gaussian. The solution is given in [9] as

$$\Omega = \sum_{p=1}^{p} \left[\frac{\delta h_p}{\delta \theta} \right]^T Q_p^{-1} \left[\frac{\delta h_p}{\delta \theta} \right], \tag{6a}$$

$$\xi = \sum_{p=1}^{p} \left[\frac{\delta h_p}{\delta \theta} \right]^T Q_p^{-1} [z'_p], \text{ and}$$
 (6b)

$$\Delta \theta = \Omega^{-1} \xi, \tag{6c}$$

where Ω is the covariance of the state-estimate correction $\Delta \theta$.

The only remaining task is to convert the posterior in (2) into the form required by (4) by defining the appropriate forms for z_p , $h_p(\theta)$, and Q_p^{-1} for each term. Then, we will be able to use (6a-6c) to solve for the unknown, HR image x. We first define the augmented state-vector θ to be the combination of the HR image x and the set of warping parameters $\{s_k\}$; i.e. $\theta = \begin{bmatrix} x \\ \{s_k\} \end{bmatrix}$.

The data likelihood term in (2), $\beta_k^{mn/2} exp\left(-\frac{\beta_k}{2} \|y_k - AHC_k(s_k)\|^2\right)$, is realized by setting $z_p = y_k$, $h_p(\theta) = AHC_k(s_k)x$, and $Q_p^{-1} = \beta_k I_{N \times N}$. $I_{N \times N}$ is the (N x N) identity matrix. Likewise, the prior on the warping parameter corrections, $exp\left(-\frac{1}{2}s_k^T\Omega_{sk}s_k\right)$, is realized by setting $z_p = 0$, $h_p(\theta) = s_k$, and $Q_p^{-1} = \Omega_{sk}$.

In order to handle the image prior model $Pr[x|\alpha]$ we use the method of "pseudo-measurements" [14] which have been applied in the domain of Kalman filter based tracking but is not widely applied to image processing. A pseudomeasurement is created by defining $z_p \equiv 0$; thereby, creating a probability term which is exclusively a function of the state θ . With this trick, we can now represent a large number of common image priors models in a form required by (4). For example, the non-Gaussian TV prior from (3) is transformed into a Gaussian by using a non-linear measurement function

$$h_{p}(\theta) = \begin{bmatrix} \left(\Delta h_{1}^{2} + \Delta v_{1}^{2}\right)^{1/4} \\ \vdots \\ \left(\Delta h_{MN}^{2} + \Delta v_{MN}^{2}\right)^{1/4} \end{bmatrix}, \text{ and } (7a)$$

$$Q_p^{-1} = \alpha I_{MN \ x \ MN}. \tag{7b}$$

3.2. Expectation-Maximization for the Hyper-Parameters

At this point we could stop if we were willing to let the user manually tune the values of the hyper-parameters α and $\{\beta_k\}$. However, as mentioned above, this is undesirable. Unfortunately, the hyper-parameters, which are correctly represented as Gamma distributions [1,2], can't be transformed into Gaussians through the method of pseudomeasurements. Instead, we settle for making point source estimates of the hyper-parameters by using the EM method [5].

The EM method defines an expectation (E-step) and a maximization (M-step). The IF from (6a-6c) is effectively the E-step. In the M-step, we refine a parameter λ (which can represent either α or β_k , as

$$\lambda^{new} = \arg\max_{\lambda} E_{\theta}[ln(\Pr[\theta, \{y_k\}|\lambda])].$$
(8)

Equation (8) may be solved in closed-form to produce

$$\beta_k^{new} = \frac{N}{\|y_k - B_k(s_k)x\|^2 + E_{\Delta\theta}[\Delta\theta^T V^T V \Delta\theta]} \quad \text{, and} \quad (9a)$$

$$\alpha^{new} = \frac{MN}{\left\|h_p(\theta_0)\right\|^2 + E_{\Delta\theta}\left[\Delta\theta^T V^T V \Delta\theta\right]} \quad , \tag{9b}$$

where $V = -AH \left[C_k(s_k) \quad \frac{\partial C_k(s_k)}{\partial s_{k1}} x \quad \dots \quad \frac{\partial C_k(s_k)}{\partial s_{kP}} x \right]$, which is $\left[\frac{\delta h_p}{\delta \theta} \right]$. The form of $h_p(\theta)$ depends upon the choice of

is $\left[\frac{r}{\delta\theta}\right]$. The form of $h_p(\theta)$ depends upon the choice of image prior model.

4. COMPARISON OF IFEM AND VBI

In order to compare the IFEM to VBI solutions, we make a comparison of the two SR methods applied to the outdoor scene shown in Figure 1. The scene was imaged from an altitude of 20m using a DJI Phantom 3 quadcopter drone with an attached 4K RGB camera. A drone is able to naturally provide the frame-to-frame motion required by multi-frame SR algorithms. As a control, we apply the non-SR BiCubic up-sampling method followed by blur deconvolution using the Richardson-Lucy algorithm [15]. We then apply both the VBI algorithm from [1] (which is available as a Matlab software package [16]) as well as our IFEM algorithm. For detailed comparison, we focus on the car license plate as well as the pair of Siemens star resolution targets (see Figure 2).

The star targets, which are recommended by the ISO 12233 standard [17], allow us to verify, quantitatively, that the two SR algorithms are actually performing the fundamental requirement of an SR algorithm; i.e., to increase resolution [18]. Resolution is defined by the ISO standard in terms of modulation transfer function (MTF) at high spatial frequencies. Figure 3 shows that both the VBI and IFEM algorithms provide similar MTF gains in the aliased region of the LR images (spatial frequencies greater than 0.5

cycles/pixel) whereas the non-SR, BiCubic algorithm does not.

5. CONCLUSION

In this paper, we have derived a solution to the SR problem which provides the statistical advantages of a full Bayesian solution with reduced derivation complexity. We believe the reduced complexity will make the advantages of the statistical solution more accessible to the general practitioner than more difficult analytic methods, such as VBI. Using sample imagery, we were able to directly compare our solution to that obtained by a state-of-the-art VBI SR algorithm.



Figure 1: Outdoor scene used to evaluate alternate SR algorithms (viewed from 20m altitude with a DJI Phantom 3







Figure 3: Comparison of MTF from Siemens star targets

6. REFERENCES

[1] S. Villena, M. Vega, R. Molina, and A. K. Katsaggelos, "A Non-stationary Image Prior Combination in Super-resolution," *Digital Signal Processing*, Vol 32, pp. 1-10, 2014.

[2] S. Babacan, R. Molina, and A. K. Katsaggelos, "Variational Bayesian Super-Resolution Reconstruction," in *Super-Resolution Imaging*, Milanfar ed., CRC Press, pp. 285-313, 2010.

[3] S. Farsiu, M. Robinson, M. Elad, and P. Milanfar, "Fast and Robust Multiframe Super Resolution," *IEEE Trans. on Image Proc.*, Vol 13, pp. 1327-1344, 2004.

[4] H. Zhang, Y. Zhang, H. Li, and T. Huang, "Generative Bayesian Image Super-Resolution with Natural Image Prior," *IEEE Trans. on Image Proc.*, Vol **21**, pp. 4054-4067, 2012.

[5] C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[6] S. Villena, M. Vega, S. Babacan, . R. Molina, and A. K. Katsaggelos, "Bayesian Combination of Sparse and Non-sparse Priors in Image Super Resolution," *Digital Signal Processing*, Vol 23, pp. 530-541, 2013.

[7] R. Ranganath, S. Gerrish, and D. Blei, "Black Box Variational Inference," in *Proc. of AISTAS*, pp. 814-822, 2014.

[8] A. Kucukelbir, R. Ranganath, A. Gelman, and D. Blei, "Automatic Variational Inference in Stan," *Proceedings of the 17th International Conference on AISTATS*, 2014.

[9] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics* Academic, 2006.

[10] F. Dellaert, C. Thorpe, and S. Thrun, "Super-Resolved Texture Tracking of Planar Surface Patches," *IEEE Proc. International Conf. on Intelligent Robots and Systems*, pp. 197-203, 1998.

[11] NVIDIA CUDA Sparse Matrix library (cuSPARSE), https://developer.nvidia.com/cusparse.

[12] E. Wan and R. van der Merwe, "The Unscented Kalman Filter for Nonlinear Estimation," in *Proc. of AS-SPCC*, pp. 153-158, 2000.

[13] G. Subrahmanyam, A. Rajagopalan, and R. Aravind, "Unscented Kalman Filter for Image Estimation in Film-Gain Noise," in *Proc. of ICIP*, pp. 17-20, 2007. [14] S. Julier and J. LaViola, "On Kalman Filtering with Nonlinear Equality Constraints," *IEEE Trans. on Signal Proc.* Vol. **55**, pp. 2774-2784, 2007.

[15] Biggs, D.S.C. "Acceleration of Iterative Image Restoration Algorithms." *Applied Optics*, Vol. 36. Number 8, pp. 1766–1775, 1997.

[16] Variational Bayesian Inference Super-Resolution Matlab Software, http://decsai.ugr.es/pi/superresolution/software.html.

[17] "Photography – Electronic still picture imaging – Resolution and spatial frequency response," ISO 12233, 2014.

[18] M. Woods and A. K. Katsaggelos, "Spatial-frequency-based Metric for Image Superresolution," *J. Opt. Soc. Am. A*, Vol 32, pp. 2002-2020, 2015.