Bayesian Image Super-Resolution

Mahdi Ranjbar, Aissa Abdelaziz
Department of Mathematics and Computer Science
ENS Paris-Saclay
{mahdi.ranjbar, aissa.abdelaziz}@ens-paris-saclay.fr

Abstract

This report examines a multi-frame image super-resolution approach from a Bayesian point of view in which the likelihood function for the image registration parameters is based on a marginalization over the unknown high-resolution image proposed by Tipping and Bishop. This approach estimates the unknown point spread function, and is rendered tractable through the introduction of a Gaussian process prior over images. The accompanying source code for this project is available for reference [1].

1 Introduction

The problem of fusing information from several low-resolution images into a single clearer high-resolution image is referred to as multi-frame super-resolution. The output image has a higher spatial resolution, or includes more visible detail in the high spatial frequencies [2]. Super-resolution can be used to improve the aesthetic quality of images for media publication, or to feed into higher-level vision tasks such as object recognition or localization [1].

The low-resolution images taken as inputs can come from a wide variety of sources, from television broadcasts and holiday snaps to surveillance footage and satellite terrain imagery. These sets of low-resolution images are related to the high-resolution version of the scene by imaging parameters, including the point-spread function, lighting variations, and geometric warpings. In order to perform super- resolution by exploiting the relative sub-pixel motion between the scene and the imaging planes, we need to estimate these imaging parameters, but no matter how accurate the estimate, there will be some residual uncertainty associated with the values [4].

1.1 Background

This basic problem of high resolution image recovery using multiple frames was first addressed in the literature by Tsai and Huang [8]. Another approach to the high resolution image reconstruction problem uses a projection onto convex sets (POCS) algorithm to solve iteratively the super-resolution inverse problem using a full generative image model and arbitrary motion model [7]. The problem has also been approached from a statistical estimation framework. Specifically, a maximum a posteriori (MAP) estimator is developed in [6]. This MAP estimator uses an edge preserving Huber-Markov random field for the image prior.

Tipping and Bishop's approach differs from previous methods by using Bayesian techniques instead of simple MAP. This allows to determine low resolution image registration parameters by marginalizing over the unknown high resolution image.

[1] https://github.com/aissa-abdelaziz/Bayesian-Image-Super-Resolution.git

1.2 Image Registration

Essential to the work on super-resolution estimation is the need to find an accurate point-to-point correspondence between images in the input sequence. This is known as registering the images. Under certain conditions the mapping can be expressed as a simple geometric relationship between the images.

1.3 The Point Spread Function

The trickiest aspect of many super-resolution systems is determining the point-spread function (PSF). It's vital as it outlines how each x pixel affects pixels in the observed images. Typically, the PSF, arising from optical blur, sensor artifacts, and possible motion during exposure, is modeled as either an isotropic Gaussian or a uniform disk in super-resolution.

The proposed Bayesian marginalization method is able to to determine the PSF and the registration parameters accurately by marginalizing over the unknown high resolution image, and thereby reconstruct the high resolution image with subjectively very good quality.

2 Bayesian Super-Resolution

2.1 Generative Model

The generative model for multi-frame super-resolution assumes that a known high-resolution image generates each low-resolution image according to pre-determined geometric transforms, and a given PSF kernel. Formally, the high-resolution image $\mathbf{x}$ (vectorized, size $N \times 1$) generates K low-resolution images, each of which is $\mathbf{y}^{(k)}$ (vectorized, size $M \times 1$). The set of K such images, $\left\{\mathbf{y}^{(k)}\right\}$ is generated from geometric registration parameters $\left\{\mathbf{s}^{(k)}, \theta^{(k)}\right\}$, representing 2D shift vectors and rotation angles, respectively. A common blur parameter γ, corresponding to the standard deviation of a Gaussian blur kernel, is used for all the images in the sequence. Gaussian i.i.d. noise with precision (inverse variance) β is then added to each $\mathbf{y}^{(k)}$,

$$\mathbf{y}^{(k)} = \mathbf{W}\left(\mathbf{s}^{(k)}, \theta^{(k)}, \gamma\right)\mathbf{x} + \mathcal{N}\left(\mathbf{0}, \beta^{-1}\mathbf{I}\right) \tag{1}$$

Each row of $\mathbf{W}^{(k)}$ constructs a single pixel in $\mathbf{y}^{(k)}$. The transformation matrix $\mathbf{W}^{(k)}$ in Equation 1 is given by a PSF which captures the down-sampling process and has a Gaussian form

$$W_{ji}^{(k)} = \widetilde{W}_{ji}^{(k)} / \sum_{i'} \widetilde{W}_{ji'}^{(k)} \quad \text{, with} \quad \widetilde{W}_{ji}^{(k)} = \exp\left\{-\frac{\left\|\mathbf{v}_i - \mathbf{u}_j^{(k)}\right\|^2}{\gamma^2}\right\} \tag{2}$$

where γ represents the width of the PSF. The vector $\mathbf{v_i}$ denotes the spatial position in the 2-dimensional image space of pixel i with $i = 1, \ldots, N$. The vector $\mathbf{u}_j^{(k)}$ is the centre of the PSF with $j = 1, \ldots M$, and is dependent on the shift and rotation of the low resolution image. A parameterization is chosen such that the centre of rotation coincides with the centre $\overline{\mathbf{v}}$ of the image, so that

$$\mathbf{u}_j^{(k)} = \mathbf{R}^{(k)}\left(\mathbf{v}_j - \overline{\mathbf{v}}\right) + \overline{\mathbf{v}} + \mathbf{s}_k \tag{3}$$

where $\mathbf{R}^{(k)}$ is the rotation matrix

$$\mathbf{R}^{(k)} = \begin{pmatrix} \cos\theta_k & \sin\theta_k \\ -\sin\theta_k & \cos\theta_k \end{pmatrix} \tag{4}$$

2.2 Likelihood and Posterior

For a set of low-resolution images assuming that they are generated independently from the model, given their registrations, the PSF, and the high-resolution image $\mathbf{x}$, the data likelihood is

$$
\begin{aligned}
\mathcal{L} &= p\left(\left\{\mathbf{y}^{(k)}\right\} \mid \mathbf{x}, \left\{\boldsymbol{s}^{(k)}, \boldsymbol{\theta}^{(k)}\right\}, \gamma\right) \\
&= \prod_{k=1}^{K} p\left(\mathbf{y}^{(k)} \mid \mathbf{x}, \boldsymbol{s}^{(k)}, \boldsymbol{\theta}^{(k)}, \gamma\right) \\
&= \prod_{k=1}^{K} \left(\frac{\beta}{2\pi}\right)^{\frac{M}{2}} \exp\left\{-\frac{\beta}{2}\left\|\mathbf{y}^{(k)} - \mathbf{W}\left(\mathbf{s}^{(k)}, \theta^{(k)}, \gamma\right)\mathbf{x}\right\|_2^2\right\}
\end{aligned}
\tag{5}
$$

Optimizing this likelihood with respect to the pixels in $\mathbf{x}$ gives the Maximum Likelihood (ML) solution. The system is ill-conditioned, however, in general a prior distribution $p(x)$ over super-resolution image is required.

$$
\begin{aligned}
\log \mathcal{L} &= \sum_{k=1}^{K} \log p\left(y^{(k)} \mid x, s^{(k)}, \theta^{(k)}, \gamma\right) \\
&= \frac{KM}{2} \log\left(\frac{\beta}{2\pi}\right) - \frac{\beta}{2} \sum_{k=1}^{K} \left\|y^{(k)} - W^{(k)} x\right\|^2
\end{aligned}
\tag{6}
$$

The posterior distribution over the high resolution image is in the form

$$
\begin{aligned}
p\left(\mathbf{x} \mid \left\{\mathbf{y}^{(k)}, \mathbf{s}_k, \theta_k\right\}, \gamma\right) &= \frac{p(\mathbf{x}) \prod_{k=1}^{K} p\left(\mathbf{y}^{(k)} \mid \mathbf{x}, \mathbf{s}_k, \theta_k, \gamma\right)}{p\left(\left\{\mathbf{y}^{(k)}\right\} \mid \left\{\mathbf{s}_k, \theta_k\right\}, \gamma\right)} \\
&= \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})
\end{aligned}
\tag{7}
$$

with

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \left[\mathbf{Z}_x^{-1} + \beta\left(\sum_{k=1}^{K} \mathbf{W^{(k)}}^{\mathrm{T}} \mathbf{W}^{(k)}\right)\right]^{-1}, \\
\boldsymbol{\mu} &= \beta \boldsymbol{\Sigma}\left(\sum_{k=1}^{K} \mathbf{W^{(k)}}^{\mathrm{T}} \mathbf{y}^{(k)}\right).
\end{aligned}
\tag{8}
$$

Thus the posterior distribution over the high resolution image is again a Gaussian process.

2.3 Marginalization

By marginalizing over $\mathbf{x}$, an ML estimate to be maximized with respect to parameters $\boldsymbol{s}, \boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ is in form

$$
\begin{aligned}
& p\left(\left\{\mathbf{y}^{(k)}\right\} \mid \left\{\boldsymbol{s}^{(k)}, \boldsymbol{\theta}^{(k)}\right\}, \gamma\right) \\
&= \int p(\mathbf{x}) p\left(\left\{\mathbf{y}^{(k)}\right\} \mid \mathbf{x}, \left\{\boldsymbol{s}^{(k)}, \boldsymbol{\theta}^{(k)}\right\}, \gamma\right) \mathrm{d}\mathbf{x} \\
&= \int p(\mathbf{x}) \prod_{k=1}^{K} p\left(\mathbf{y}^{(k)} \mid \mathbf{x}, \boldsymbol{s}^{(k)}, \boldsymbol{\theta}^{(k)}, \gamma\right) \mathrm{d}\mathbf{x}
\end{aligned}
\tag{9}
$$

This integral is generally only tractable if a Gaussian form is chosen for the image prior $p(x)$.

2.4 Gaussian Prior

$$
\begin{aligned}
p(x) &= \mathcal{N}\left(\mathbf{x} \mid \mathbf{0}, \mathbf{Z}_x\right) \\
&= \frac{1}{\left(2\pi \det\left(Z_x\right)\right)^{\frac{N}{2}}} \exp\left(-\frac{1}{2} x^T Z_x^{-1} x\right)
\end{aligned}
\tag{10}
$$

Where $Z_x \in \mathbb{R}^{N \times N}$, and

$$Z_x(i,j) = A \cdot \exp\left(-\frac{\|v_i - v_j\|^2}{r^2}\right)$$

A, r are constants that model the prior strength and the correlation length scale respectively. These are fixed, and are not estimated. Thus, Z_x remains constant throughout, and thus can be pre-computed and re-used.

2.5 Marginalizing over the Super-Resolution Image

We can substitute Equation 10, and Gaussian observation density of Equation 5 into Equation 9, which marginalizes over the super-resolution image.

Taking y and W to be the vector and matrix of stacked $\mathbf{y}^{(k)}$ and $\mathbf{W}^{(k)}$ to be a stacked vector of all the input images, this distribution is

$$p\left(\mathrm{y} \mid \{\mathbf{s}_k, \theta_k\}, \gamma\right) = \mathcal{N}\left(\mathbf{0}, \mathbf{Z}_y\right) \tag{11}$$

where

$$\mathbf{Z}_y = \beta^{-1}\mathbf{I} + \mathbf{W}\mathbf{Z}_x\mathbf{W}^{\mathrm{T}},$$

Using some standard matrix manipulations we can rewrite the marginal likelihood in the form

$$\begin{aligned}\log p\left(\mathbf{y} \mid \{\mathbf{s}_k, \theta_k\}, \gamma\right) = -\tfrac{1}{2}\Big[\beta \textstyle\sum_{k=1}^{K} \left\|\mathbf{y}^{(k)} - \mathbf{W}^{(k)}\boldsymbol{\mu}\right\|^2 + \boldsymbol{\mu}^{\mathrm{T}}\mathbf{Z}_x^{-1}\boldsymbol{\mu} \\ + \log|\mathbf{Z}_x| - \log|\boldsymbol{\Sigma}| - KM\log\beta\Big]\end{aligned} \tag{12}$$

We now optimize this marginal likelihood with respect to the parameters $\left\{\mathbf{s}^{(k)}, \theta^{(k)}, \right\}, \gamma$.

2.6 Implementation

In implementing the algorithm, we first generate K synthetic low-resolution images. These images are created through random shifts, rotations, and additive noise (with a given sigma value). In the original approach by Tipping et al., they use Scaled Conjugate Gradients for minimization. In our case, however, we utilize the Adam optimizer along with PyTorch gradient descent provided by PyTorch. Following this step, we proceed to optimize the full posterior (as defined in Eq 7) This optimization process is divided into two steps:

Initially, we estimate the registration parameters (theta, epsilon, gamma), as detailed in the paper. Due to memory constraints, we utilize small patches extracted from the low-resolution image to optimize the marginal Log Likelihood (Eq 12) .

After obtaining the optimal registration parameters, we initialize the high-resolution image using bilinear interpolation, and then proceed to optimize the numerator of the full posterior equation over the high-resolution images using the Adam optimizer. This step aims to improve the quality of the high-resolution images.

To quantitatively evaluate the results obtained, we use Peak Signal to Noise Ratio (PSNR) and Root Mean Square Error (RMSE).

3 Experiments

We use three synthetic datasets to measure quantitatively and qualitatively the performance with respect to known ground truth high-resolution images.

Initially, we resize all high-resolution images to a standard size of $[150, 150]$, and then scale them to ensure intensity values fall within the range of $(-0.5, 0.5)$. Subsequently, we generate 16 synthetic low-resolution images using the observation model (Eq 1) with a scale factor of 4 to have size

of $[37, 37]$. We apply a shift and rotation drawn from a uniform distribution over the interval : $\mathbf{s} \in (-4, 4)$, $\theta \in (-2, 2)$ and we add some additive noise with standard deviation of $\sigma = 0.02$.

We use the Gaussian process with width parameter $r = 1.0$, variance parameter $A = 0.04$.

In order to compare with other processes, we also utilize Markov Random Field with the same hyperparameters. We use PSNR and RMSE to calculate the quantitative results as provided in Table 1. Additionally, we provide the qualitative results in Figure **??**.

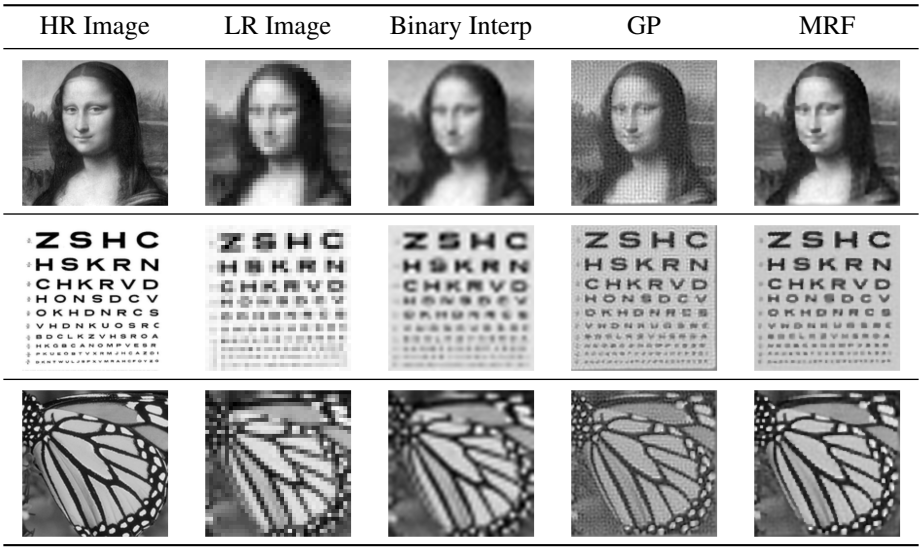
PSNR :measures the quality gap between a reconstructed image and the original, often reflecting compression artifacts or distortion.

RMSE :measures the average deviation between predicted and observed values, assessing how accurately a model fits the data.

Table 1: PSNR and RMSE Results using different Priors

Prior	PSNR (↑)	RMSE (↓)
Bilinear interpolation	12.95	0.225
Gaussian Prior	17.959	0.139
MRF Prior	20.502	0.121

Figure 1: High-resolution images generated from 16 low-resolution versions of the Monaliza, eyechart, and butterfly datasets at a scale factor of 4 using linear interpolation and Bayesian methods with different priors: Gaussian Prior (Main), and Markov Random Field. The variation in intensity is clearly visible.



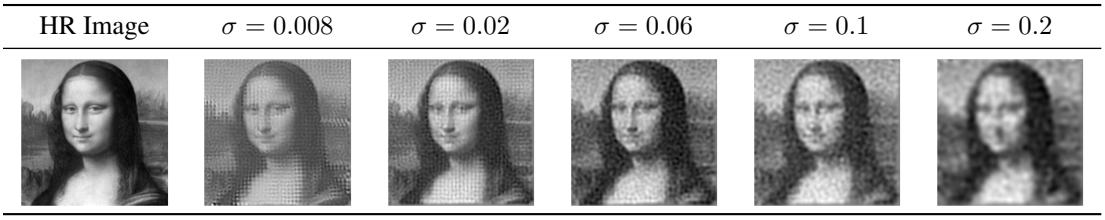
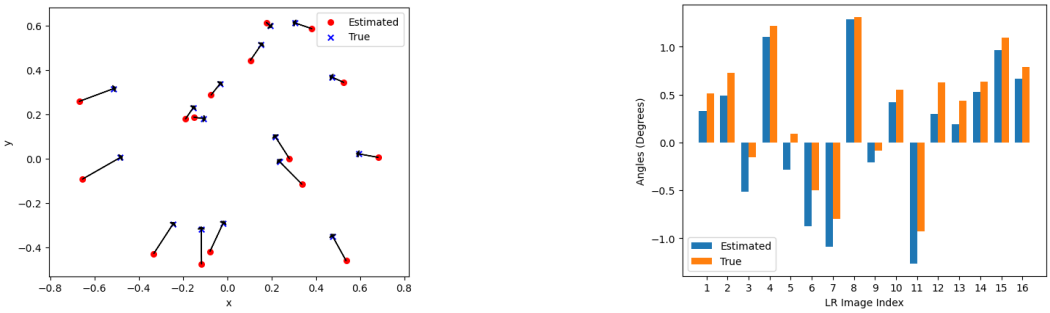


Figure 2: The HR image gets progressively worse from left to right, with the standard deviations being 0.02, 0.06, 0.10, and finally 0.2 .



(a) Estimated shifts

(b) Estimated angles

Figure 3: Plot showing the estimated shift and algles obtained after optimizing the marginal likelihood with respect to the registration parameter.

Algorithm 1 Bayesian Image Super-Resolution

Step 1: Scale the high-resolution image and set all the parameters.
Step 2: Generate a synthetic low-resolution image by applying shift, rotation, and additive Gaussian noise.
Step 3: Extract a small patch from the low-resolution image to be used for the next step.
Step 4: Optimize using Adam the marginal likelihood with respect to the registration parameters (Eq 12).
Step 5: Fix the best estimated parameters and initialize the high-resolution image using binary interpolation.
Step 6: Optimize using Adam the full posterior distribution with respect to high-resolution images (Eq 7).
Step 7: Fine-tune the model by repeating the previous step to get the best result.

4 Limitations

Tipping and Bishop's method has three limitations:

(1) it is restricted to a Gaussian image prior in order for the marginalization to remain tractable, but others have shown improved image super-resolution results are produced using distributions with heavier tails [5]. Specifically, in [3], by marginalizing over unknown registration parameters, $p(x)$ can be taken outside of the integral in Equation 9, which allows to choose any suitable super-resolution image prior.

(2) it is computationally expensive due to the very large matrices required by the algorithm. Indeed, the optimization over registration and blur parameters is carried out with low-resolution images patches of just 12x12 pixels, rather than the full low-resolution images because of the computational cost involved in computing the terms in Equation 12.

(3) it does not take into account photometric model but only motion model.

5 Discussion

Our initial findings, which involve estimating image registration and PSF parameters by marginalizing likelihood over the unknown super-resolution image with a Gaussian prior (see Figure 3), exhibit commendable accuracy. However, the computational complexity remains a challenge that requires attention.

We observed noise amplification, particularly at edges, and some loss of detail in artifacts post-reconstruction. This phenomenon is attributed to the heavy reliance of edge pixels on the Gaussian prior, owing to limited evidence from low-resolution images.

Furthermore, varying the noise standard deviation (σ) reveals its significant impact on reconstruction quality. Higher noise levels correlate with inferior outcomes, underlining the method's practical importance.

Repeated executions of the algorithm highlight the critical role of precise parameter selection for achieving satisfactory results.

Pickup et al. [3], proposed marginalizing over the unknown registration parameters relating the set of input low-resolution views instead of marginalizing over the super-resolution image. This results in more realistic prior distributions, and also reduces the dimension of the integral considerably, removing the main computational bottleneck of the other algorithm. They also introduced the illumination components into the generative model to handle changes in lighting as well as motion.

6 Conclusion

In this work, we explored the 'Multi-Frame Super-Resolution' problem and its Bayesian formulation, which involves marginalization over the unknown high-resolution image using a Gaussian process prior. We derived the mathematical framework, conducted various experiments, and compared the results with a different prior.

References

[1] Russell C Hardie, Kenneth J Barnard, and Eyal E Armstrong. "Joint map registration and high-resolution image estimation using a sequence of undersampled images". In: *IEEE Transactions on Image Processing* 6 (1997), pp. 1621–1633.

[2] Michal Irani and Shmuel Peleg. "Super resolution from image sequences". In: *Proceedings of the International Conference on Pattern Recognition*. Vol. 2. 1990, pp. 115–120.

[3] Lyndsey C Pickup et al. "Bayesian image super-resolution, continued". In: *Advances in Neural Information Processing Systems (NIPS)*. Vol. 19. 2006.

[4] Daniel Robinson and Peyman Milanfar. "Fundamental performance limits in image registration". In: *IEEE Transactions on Image Processing* 13 (2004), pp. 1185–1199.

[5] R. R. Schultz and R. L. Stevenson. "A Bayesian approach to image expansion for improved definition". In: *IEEE Transactions on Image Processing* 3 (1994), pp. 233–242.

[6] Robert R. Schultz and Richard L. Stevenson. "Extraction of high-resolution frames from video sequences". In: *IEEE Transactions on Image Processing* 5 (June 1996), pp. 996–1011.

[7] Henry Stark and Parviz Oskoui. "High-resolution image recovery from image-plane arrays, using convex projections". In: *Journal of the Optical Society of America A* 6.11 (1989), pp. 1715–1726.

[8] Roger Tsai and Thomas Huang. "Multiframe image restoration and registration". In: *Advances in Computer Vision and Image Processing*. Vol. 1. JAI Press Inc., 1984, pp. 317–339.