

HANDLING EXPOSURE TIME IN MULTI-FRAME IMAGE RESTORATION

Bahadir K. Gunturk

Electrical and Computer Engineering
Louisiana State University
Baton Rouge, LA 70803
email: bahadir@ece.lsu.edu

ABSTRACT

The process of reconstructing a high-resolution image from multiple low-resolution images is known as super-resolution reconstruction. One of the assumptions of current super-resolution reconstruction methods is that all images are captured with the same exposure time and aperture size. This is not necessarily true. In reality, cameras have limited dynamic range and nonlinear response to the quantity of light received; and camera settings might be adjusted automatically or manually to capture the desired portion of the scene's dynamic range. In this paper, we propose a super-resolution algorithm based on an imaging model that includes camera response function, exposure time, sensor noise and quantization error in addition to spatial blurring and sampling. The algorithm is based on Bayesian estimation. Initial experiments demonstrate the effectiveness of the algorithm.

1. INTRODUCTION

With the development of visual communications and image processing applications, there is a high demand for high-resolution images not only to give the viewer a high-quality picture but also to provide additional detail that may be critical in various applications. Digital cameras, surveillance systems, medical imaging, aerial/satellite imaging, and high-definition TV systems are some of the application areas where high-resolution images are desired. For example, in medical imaging high-resolution images are required to make correct diagnosis and operational decisions. Surveillance systems require high-resolution images to recognize faces, licence plates, etc. In aerial/satellite imaging, high-resolution images are required to resolve small objects and to make correct detection/classification decisions.

The most direct way of increasing spatial resolution is to increase the number of sensor elements per unit area. Although this can be achieved by reducing pixel size and placing pixels more densely, the cost of producing such sensor arrays may not be appropriate for general purpose commercial applications. More importantly, as pixel size decreases, the image quality degrades because of shot noise. Shot noise is due to the inherent quantum uncertainty in the electron-hole pair generation process and remains roughly the same with reduced pixel size, whereas the signal power decreases proportional to the pixel size reduction.

An alternative approach is to use signal processing techniques to improve spatial resolution. When there are multiple images of a scene, it is possible to increase the spatial resolution by exploiting the correlation among those images. Such a multi-frame resolution

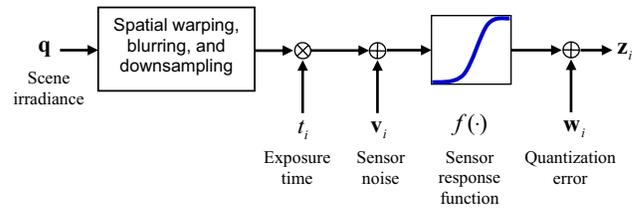


Fig. 1. The proposed super-resolution algorithm uses an imaging model that includes dynamic range and spatial domain effects.

enhancement process is referred to as super-resolution reconstruction in the literature [1, 7, 11, 8, 13].

Although considerable work has been done in the area of super-resolution reconstruction, an important drawback of these algorithms is the assumption that all images (to be used in reconstruction) capture the same portion of the dynamic range. In other words, it is assumed that camera parameters such as exposure time and aperture size are fixed for all images. In fact, sensors have limited dynamic range, and the camera parameters need to be adjusted to capture the right portion of the scene's dynamic range. All modern cameras are equipped with automatic parameter control units. Therefore, the assumption of fixed camera parameters fails unless the parameters are fixed manually, which is not desirable in video imagery because of the wide dynamic range and potential illumination changes. In addition, it is possible to obtain more information about a scene by combining images that are captured with different camera parameters.

The so-called high-dynamic range imaging has been an active research area in the computer vision community. Debevec *et al.* [4], Mann [9], Robertson *et al.* [12], and Candocia [2] have demonstrated how to improve dynamic range by combining images captured with different exposure times. However, the issue of nonlinear sensor response and different exposures has not been addressed extensively in the super-resolution research. Recently, we proposed a projections onto convex sets based super-resolution algorithm that addresses the saturation in pixel measurements and potential changes in illumination [6]. In this paper, we propose a stochastic super-resolution reconstruction algorithm that models nonlinear camera response function, exposure time, sensor noise, and quantization error in addition to spatial blurring and sampling. In Section 2, we present the image acquisition model used. In Section 3, we provide the derivation and the details of the proposed algorithm. Preliminary experimental results are given in Section 4.

2. IMAGING MODEL

Incorporating the relative motion among observed images, super-resolution algorithms model the imaging process as a linear mapping between a high-resolution input signal \mathbf{q} and low-resolution observations \mathbf{z}_i . ($i = 1, \dots, N$; N is the total number of observations.) The imaging process is formulated as

$$\mathbf{z}_i = \mathbf{H}_i \mathbf{q}, \quad i = 1, \dots, N, \quad (1)$$

where \mathbf{H}_i is the linear mapping that includes motion (of the camera or the objects in the scene), blur (caused by the point spread function of the sensor elements and the optical system), and downsampling. Therefore, super-resolution reconstruction is an inverse problem where \mathbf{q} is estimated from a set of observations \mathbf{z}_i . \mathbf{H}_i can be space- and time-varying. In practice, \mathbf{H}_i is implemented in three steps: spatial warping to compensate for motion, convolution with a point spread function (PSF), and downsampling. Details of \mathbf{H}_i modeling can be found in the special issue of the IEEE Signal Processing Magazine [1] and the references therein.

As mentioned earlier, the model in equation (1) is not a complete model. An imaging sensor has a nonlinear response to the quantity of light it receives, and only a portion of the scene's dynamic range is captured. Exposure time may not be identical for all images. In addition to the exposure time, we also need to consider sensor noise and quantization error in the imaging process. Denoting \mathbf{v}_i as the additive noise term (due to shot noise and thermal noise) and \mathbf{w}_i as the quantization error, the imaging process can be formulated as

$$\mathbf{z}_i = f(t_i \mathbf{H}_i \mathbf{q} + \mathbf{v}_i) + \mathbf{w}_i, \quad i = 1, \dots, N, \quad (2)$$

where $f(\cdot)$ is the nonlinear camera response function, t_i is the exposure time, and \mathbf{H}_i is the linear mapping that incorporates motion, PSF, and downsampling. (See the block diagram in Figure 1.)

3. RECONSTRUCTION ALGORITHM

3.1. Stochastic Estimation

In this section, we derive a maximum-likelihood estimator for \mathbf{q} . Defining $g(\cdot) \equiv f^{-1}(\cdot)$ and using a Taylor series expansion, equation (2) can be written as

$$g(\mathbf{z}_i) \simeq t_i \mathbf{H}_i \mathbf{q} + \mathbf{v}_i + g'(\mathbf{z}_i) \mathbf{w}_i. \quad (3)$$

We model \mathbf{v}_i and \mathbf{w}_i as zero mean independent identically distributed (IID) Gaussian noises with variances σ_v^2 and σ_w^2 , respectively. This will result in an analytically trackable derivation. It can be shown that the total noise, $\mathbf{v}_i + g'(\mathbf{z}_i) \mathbf{w}_i$, is also a zero mean Gaussian noise with variance

$$\sigma^2 = \sigma_v^2 + g'(\mathbf{z}_i)^2 \sigma_w^2. \quad (4)$$

A critical implication of this result is that the total noise variance σ^2 is a function of the camera response function and measured pixel intensities \mathbf{z}_i . Equation (4) indicates that the total noise variance is larger for saturated pixel values. This will become more clear when we show a typical camera response function shortly.

Denoting \mathbf{K} as the covariance matrix of the total noise, the maximum-likelihood estimate of \mathbf{q} minimizes the following cost function:

$$E(\mathbf{q}) = \sum_i \left(\frac{g(\mathbf{z}_i)}{t_i} - \mathbf{H}_i \mathbf{q} \right)^T \mathbf{K}^{-1} \left(\frac{g(\mathbf{z}_i)}{t_i} - \mathbf{H}_i \mathbf{q} \right). \quad (5)$$



Fig. 2. Images of the same scene captured with a Canon G5 digital camera. The exposure times are 1/25, 1/50, 1/100, 1/200, 1/400, 1/800, and 1/1250 seconds. Original image size is 480×640 .

Because of the IID noise assumption, \mathbf{K} is a diagonal matrix, and its diagonal is equal to σ^2 .

One technique to obtain the maximum-likelihood estimate in equation (5) is the steepest descent technique. \mathbf{q} can be estimated by iteratively updating an initial estimate in the direction of the negative gradient of $E(\mathbf{q})$. At the k th iteration, the estimate is

$$\mathbf{q}^{(k)} = \mathbf{q}^{(k-1)} - \alpha \nabla E(\mathbf{q}^{(k-1)}), \quad (6)$$

where α is the step size, and $\nabla E(\mathbf{q})$ can be found as

$$\nabla E(\mathbf{q}) = - \sum_i \frac{1}{t_i} \mathbf{H}_i^T \mathbf{K}^{-1} (g(\mathbf{z}_i) - t_i \mathbf{H}_i \mathbf{q}). \quad (7)$$

The step size α in equation (6) can be fixed or updated adaptively during the iterations. Hessian of $E(\mathbf{q})$ can be used for changing α .

As a final note, the maximum-likelihood estimation derived so far can be easily extended to maximum *a posteriori* estimation by assuming a Gaussian model (or another appropriate model) for the prior distribution of \mathbf{q} .

3.2. Complete Algorithm

In the reconstruction, everything but \mathbf{q} is either known or estimated/set in advance. The linear mapping \mathbf{H}_i requires subpixel-accurate spatial registration parameters, point spread function, and downsampling factor. Typically, the registration parameters are estimated; while the point spread function and the downsampling factor are decided in advance. Camera response function and relative exposure times can be estimated using multiple differently exposed images [4, 12, 9, 10, 15, 5].

The algorithm starts with an initial estimate $\mathbf{q}^{(0)}$, which can be obtained by (i) interpolating of one of the observations bilinearly, (ii) applying $g(\cdot)$, and (iii) dividing by the corresponding exposure time. This reference image is then updated iteratively as in equation (6). Each iteration requires simple image operations, such as warping, convolution, sampling, and scaling:

- Application of \mathbf{H}_i involves warping $\mathbf{q}^{(k)}$ to the i th frame, convolving with the point spread function, and then down-sampling.
- Camera response function is estimated in advance; therefore, calculation of $g(\mathbf{z}_i)$ is simply a look-up-table operation.
- Because \mathbf{K}^{-1} is diagonal, its application is division of each pixel value by the corresponding σ^2 .
- \mathbf{H}_i^T is implemented by upsampling the image (with zero padding), convolving with the flipped point spread function, and motion warping back to the reference frame [7].

4. EXPERIMENTAL RESULTS

4.1. Data Set

In this paper, we simulated relative motion and sampling by shifting and downsampling the images in Figure 2 by four. Four slightly shifted images of size 120×160 are picked from each differently exposed image. That is, a total of 28 images are obtained and used for restoration. Also note that none of the shifts are same among all these 28 images.

4.2. Estimating the Registration Parameters

One critical issue in super-resolution reconstruction is to find the spatial registration parameters. The problem becomes more complicated when the images are captured with different exposure times. We estimate registration parameters based on mutual information. Mutual information has been successfully used for intermodal registration [16, 3, 14]. Our experiments show that it is also effective for registering differently exposed images.

For convenience, we briefly repeat the idea of using mutual information in registration. Let $\mathbf{x} = (l_1, l_2)^T$ be the column vector containing pixel coordinates and $\mathbf{W}(\mathbf{x}; \mu)$ be a parameterized warp, where μ is a vector of parameters. $\mathbf{W}(\mathbf{x}; \mu)$ can represent affine, perspective, or any other parametric transformation. We would like to find the parameters μ that maximizes the mutual information between two images $z_i(\mathbf{W}(\mathbf{x}, \mu))$ and $z_j(\mathbf{x})$. Let $p(I_i, I_j; \mu)$ be the joint histogram of $z_i(\mathbf{W}(\mathbf{x}, \mu))$ and $z_j(\mathbf{x})$, where I_i and I_j are pixel intensities in the i th and j th images, respectively. $I_i \in [0, 255]$ and $I_j \in [0, 255]$. The marginal histograms for i th and j th images are

$$p_i(I_i; \mu) = \sum_{I_j} p(I_i, I_j; \mu) \text{ and } p_j(I_j; \mu) = \sum_{I_i} p(I_i, I_j; \mu). \quad (8)$$

Then the mutual information is

$$M(\mu) = \sum_{I_i} \sum_{I_j} p(I_i, I_j; \mu) \log_2 \left(\frac{p(I_i, I_j; \mu)}{p_i(I_i; \mu) p_j(I_j; \mu)} \right). \quad (9)$$

In our experiments, we did exhaustive search in the μ space to find the parameters that maximizes the mutual information. To obtain subpixel accurate parameters, we interpolated images using bicubic interpolation before the search. The interpolation factor was set to eight to obtain 1/8 pixel accuracy.

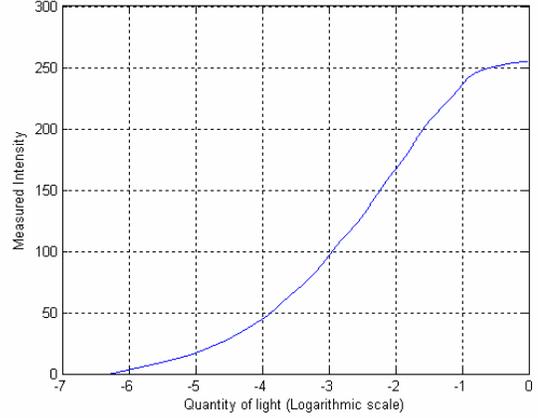


Fig. 3. Estimated camera response function is shown in logarithmic scale.

4.3. Estimating the Camera Response Function

There are various methods available in the literature to estimate camera response function [4, 12, 9, 10, 15, 5]. In our experiments, we use the comparagram method in [10]. Two differently exposed images were used to estimate the camera response function. The camera response function can be determined up to a scale factor; therefore, $f(255)$ is set to one. To obtain a smooth function, a regularization filter of $[1, -2, 1]$ is used. Estimated camera response function is shown in Figure 3. This response function is used for the rest of the experiments. When exposure times are not known, sensor response function and exposure times could be estimated iteratively [10]. In our experiments, the exposure times were known from the camera settings; therefore, iterative estimation was not necessary.

4.4. Reconstruction

In Figure 3 (linear scale), the pixel measurements that are close to 255 correspond to the saturated part of the camera response function. Therefore, the noise variance is larger for those values. After finding $g'(\cdot)$ numerically (first order difference), we obtain the inverse covariance matrix \mathbf{K}^{-1} as a function of pixel values. (The specific choices of σ_v^2 and σ_w^2 are 0.001 and 1, respectively. \mathbf{K}^{-1} is scaled such that its maximum value is one.) The initial estimate is obtained by bilinearly upsampling one of the 1/1250 second images by three, and applying $g(\cdot)$. This initial estimate is updated iteratively using equation (6). The step size α is set to 0.4; and the number of iterations is set to 15. The point spread function is chosen to be a 9×9 Gaussian window with standard deviation of 1.5.

4.5. Results

Figures 4 and 5 show two regions from bilinearly interpolated input images (one from each exposure time) and the restored image. These regions are selected from inside and outside the scene. Comparison of input data and reconstructed image demonstrates spatial resolution and dynamic range enhancement. The advantage of the stochastic approach is that noise statistics are included in the formulation in an explicit way. Prior image models can also be included easily in the stochastic approach.



Fig. 4. First seven images are seven of the bilinearly interpolated observations. Last image is the restored image.

5. CONCLUSIONS

We proposed a super-resolution algorithm that can handle changes in exposure time, and demonstrated the idea with a set of images. Unlike the previous super-resolution algorithms, the proposed algorithm uses an imaging model that includes nonlinear camera response function. The preliminary results are promising. Some of the parameters used in the reconstruction were chosen heuristically. As a future work, we will investigate estimating optimal parameters and we will include prior models in the reconstruction. We also plan to test and compare the algorithms with various other data sets.

6. REFERENCES

- [1] "Super-resolution image reconstruction," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21–86, May 2003.
- [2] F. M. Candocia, "A least squares approach for the joint domain and range registration of images," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 4, May 2002, pp. 3237–3240.
- [3] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal, "Automated multi-modality image registration based on information theory," *Information Processing in Medical Imaging*, pp. 263–274, 1995.
- [4] P. E. Debevec and J. Malik, "Modeling and rendering architecture from photographs," in *Proc. of the ACM SIGGRAPH*, 1997, pp. 369–378.
- [5] M. D. Grossberg and S. K. Nayar, "Determining the camera response from images: what is knowable?" *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 11, pp. 1455–1467, November 2003.

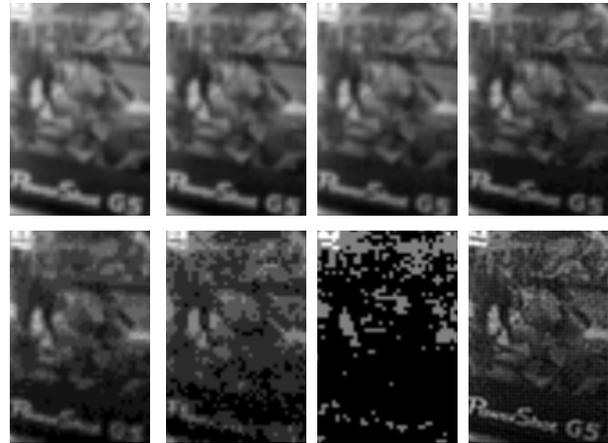


Fig. 5. First seven images are seven of the bilinearly interpolated observations. Last image is the restored image.

- [6] B. K. Gunturk, Y. Altunbasak, and R. M. Mersereau, "Multi-frame information fusion for gray-scale and spatial enhancement of images," in *Proc. IEEE Int. Conf. Image Processing*, vol. 2, September 2003, pp. 319–322.
- [7] —, "Super-resolution reconstruction of compressed video using transform-domain statistics," *IEEE Trans. Image Processing*, vol. 13, no. 1, pp. 33–43, January 2004.
- [8] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical Models and Image Processing*, vol. 53, pp. 231–239, May 1991.
- [9] S. Mann, "Comparametric equations with practical applications in quantigraphic image processing," *IEEE Trans. Image Processing*, vol. 9, no. 8, pp. 1389–1406, August 2000.
- [10] S. Mann and R. Mann, "Quantigraphic imaging: estimating the camera response and exposures from differently exposed images," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, vol. 1, 2001, pp. 842–849.
- [11] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time," *IEEE Trans. Image Processing*, vol. 6, no. 8, pp. 1064–1076, August 1997.
- [12] M. A. Robertson, S. Borman, and R. L. Stevenson, "Dynamic range improvement through multiple exposures," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, 1999, pp. 159–163.
- [13] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Processing*, vol. 5, no. 6, pp. 996–1011, June 1996.
- [14] P. Thevenaz and M. Unser, "Optimization of mutual information for multiresolution image registration," *IEEE Trans. Image Processing*, vol. 9, no. 12, pp. 2083–2099, December 2000.
- [15] Y. Tsin, V. Ramesh, and T. Kanade, "Statistical calibration of the ccd imaging process," in *Proc. Int. Conf. Computer Vision*, vol. 1, 2001, pp. 480–487.
- [16] P. Viola and W. M. Wells, "Alignment by maximization of mutual information," in *Proc. of Int. Conf. on Computer Vision*, 1995, pp. 16–23.