High-Resolution Image Reconstruction From Multiple Differently Exposed Images

Bahadir K. Gunturk, *Member, IEEE,* and Murat Gevrekci, *Student Member, IEEE*

***Abstract*—Super-resolution reconstruction is the process of reconstructing a high-resolution image from multiple low-resolution images. Most super-resolution reconstruction methods assume that exposure time is fixed for all observations, which is not necessarily true. In reality, cameras have limited dynamic range and nonlinear response to the quantity of light received, and exposure time might be adjusted automatically or manually to capture the desired portion of the scene's dynamic range. In this letter, we propose a Bayesian 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.**

***Index Terms*—Bayesian estimation, high-dynamic range imaging, multi-frame image reconstruction, super-resolution.**

I. Introduction

HIGH-RESOLUTION images are demanded 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.

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. Also, as pixel size decreases, the image quality may degrade because of shot noise.

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 enhancement process is referred to as super-resolution reconstruction in the literature [1]. Super-resolution algorithms model the imaging process between an unknown high-resolution image and multiple low-resolution observations and try to solve the inverse problem.

Manuscript received August 5, 2005; revised November 2, 2005. This work was supported by the National Science Foundation under Grant ECS-0528785. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Maja Bystrom.

The authors are with the Electrical and Computer Engineering Department, Louisiana State University, Baton Rouge, LA 70803 USA (e-mail: bahadir@ece.lsu.edu; lgevre1@ece.lsu.edu).

Digital Object Identifier 10.1109/LSP.2005.863693

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. References [2]–[5] 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 super-resolution research. Recently, Capel and Zisserman [6] used a linear model for radiometric changes, where global gain and offset parameters are estimated. This is a good model when the brightness changes are small and there is no saturation. In this letter, 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. The algorithm works even if there is saturation.

II. 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 $\mathbf{z}_i$. ($i = 1, \ldots, N$; N is the total number of observations.) The imaging process is formulated as

$$\mathbf{z}_i = \mathbf{H}_i \mathbf{q}, \qquad i = 1, \ldots, N \tag{1}$$

where $\mathbf{H}_i$ is the linear mapping that includes motion (of the camera and the objects in the scene), blur (caused by the point spread function of the sensor elements and the optical system), and downsampling [7]. 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

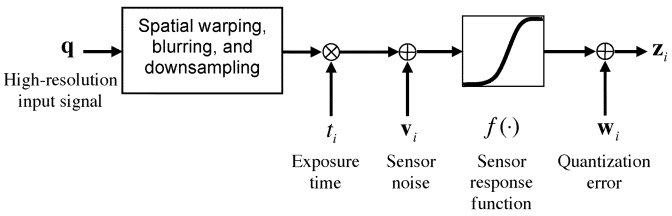

Fig. 1. Proposed super-resolution algorithm uses an imaging model that includes dynamic range and spatial domain effects.

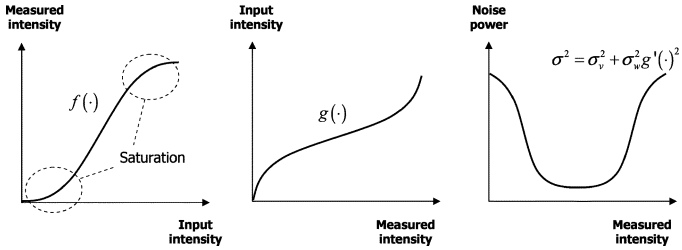

Fig. 2. For a typical sensor response function shown on the left, the total noise variance has a shape similar to the one on the right. For saturated regions, the noise variance is larger.

downsampling. Details of $\mathbf{H}_i$ modeling can be found in [1], [6], and [7].

As mentioned earlier, 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 and $\mathbf{w}_i$ as the quantization error, the imaging process can be formulated as

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

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 Fig. 1).

III. Reconstruction Algorithm

A. Derivation

Defining $g(\cdot) \equiv f^{-1}(\cdot)$ and using a Taylor series expansion, (2) can be written as

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

where "$\times$" indicates element-by-element multiplication of two vectors. Note that this derivation assumes that $f^{-1}(\cdot)$ exists. It is possible to guarantee this with parametric modeling of the camera response function [5], [8] or defining it as a one-to-one mapping [2]. Also, the approximation above neglects the second- and higher-order terms. This is a good approximation when the quantization error $\mathbf{w}_i$ is relatively small compared to $\mathbf{z}_i$.

We model the sensor noise and quantization error as independent random variables with Gaussian distributions. Such modeling has also been used previously [9]. We assume that there is no correlation among the noise values at different pixel locations. As a result, we will have an analytically tractable solution. The means of the sensor noise and quantization error are assumed to be zero. Now, let σ_v^2 and σ_w^2 be the variances of the sensor noise and quantization error. It can be shown that the total noise $v_i + g'(z_i)w_i$ is also a zero-mean independent identically distributed (i.i.d.) Gaussian random variable with variance

$$\sigma^2(z_i) = \sigma_v^2 + g'(z_i)^2\sigma_w^2 \tag{4}$$

for all pixel locations in the image $\mathbf{z}_i$. Note that the total noise variance $\sigma^2(z_i)$ is a function of the camera response function and measured pixel intensities z_i. Equation (4) indicates that the total noise variance is larger for saturated pixel values. This is illustrated in Fig. 2.

Denoting $\mathbf{K}$ as the covariance matrix of the total noise, and using a Gaussian prior for $\mathbf{q}$ with mean image $\boldsymbol{\mu}_\mathbf{q}$ and covariance matrix $\boldsymbol{\Lambda}$, the maximum a posteriori (MAP) estimate of $\mathbf{q}$ minimizes the following cost function:

$$E(\mathbf{q}) = \frac{1}{2}\sum_i \left(\frac{g(\mathbf{z}_i)}{t_i} - \mathbf{H}_i\mathbf{q}\right)^T \mathbf{K}^{-1}\left(\frac{g(\mathbf{z}_i)}{t_i} - \mathbf{H}_i\mathbf{q}\right) + \frac{1}{2}(\mathbf{q} - \boldsymbol{\mu}_\mathbf{q})^T\boldsymbol{\Lambda}^{-1}(\mathbf{q} - \boldsymbol{\mu}_\mathbf{q}). \tag{5}$$

Because of the i.i.d. noise assumption, $\mathbf{K}$ is a diagonal matrix, and its diagonal is equal to $\sigma^2(\mathbf{z}_i)$. In our experiments, we will assume that there is no spatial correlation among pixels in $\mathbf{q}$; therefore, $\boldsymbol{\Lambda}$ will be taken as a constant diagonal matrix. This simple prior serves as a regularization term. The mean image is typically obtained by interpolating and averaging the observations.

Gradient descent techniques can be used to solve (5). $\mathbf{q}$ can be estimated by iteratively updating an initial estimate in the direction of the negative gradient of $E(\mathbf{q})$. At the kth iteration, the estimate is

$$\mathbf{q}^{(k)} = \mathbf{q}^{(k-1)} - \alpha\nabla E\left(\mathbf{q}^{(k-1)}\right) \tag{6}$$

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

$$\nabla E(\mathbf{q}) = -\sum_i \mathbf{H}_i^T\mathbf{K}^{-1}\left(\frac{g(\mathbf{z}_i)}{t_i} - \mathbf{H}_i\mathbf{q}\right) + \boldsymbol{\Lambda}^{-1}(\mathbf{q} - \boldsymbol{\mu}_\mathbf{q}). \tag{7}$$

The step size α in (6) can be fixed or updated adaptively during the iterations. The Hessian of $E(\mathbf{q})$ can be used for changing α.[1]

B. 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 [2], [4], [8], [10], [11].

[1]In our experiments, α is updated at each iteration using the formula

$$\alpha = \frac{\left(\nabla E\left(\mathbf{q}^{(k-1)}\right)\right)^T\left(\nabla E\left(\mathbf{q}^{(k-1)}\right)\right)}{\left(\nabla E\left(\mathbf{q}^{(k-1)}\right)\right)^T H\left(\nabla E\left(\mathbf{q}^{(k-1)}\right)\right)}$$

where H is the Hessian matrix found by

$$H = \sum_i \mathbf{H}_i^T\mathbf{K}^{-1}\mathbf{H}_i + \boldsymbol{\Lambda}^{-1}.$$

Fig. 3. 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 s.

The algorithm starts with an initial estimate $\mathbf{q}^{(0)}$, which can be obtained by 1) interpolating of one of the observations bilinearly; 2) applying $g(\cdot)$; and 3) dividing by the corresponding exposure time. This reference image is then updated iteratively, as in (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 ith frame, convolving with the point spread function, and then downsampling [12].
- 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 [12].

Note that the Hessian-based step-size α can also be calculated without constructing large matrices [7].

IV. Experimental Results

A. Data Set

In this letter, we provide experiments with two data sets. The first data set is a half-synthetic data set. We captured seven images with different exposure times and then simulated relative motion by shifting and downsampling them. From each image, four slightly shifted images are obtained. That is, 28 images in total are obtained and used for restoration. The original seven images are shown in Fig. 3. The second data set is a real data set. We captured 15 images with a handheld digital camera. Exposure times were manually changed for some of the images. Three of these 15 images are shown in Fig. 4.

B. Estimating the Registration Parameters and Exposure Time

In our experiments, spatial registration parameters are first estimated using the feature-based method explained in [6]. The method works as follows. First, feature points in the images are extracted using the Harris corner detector [13]. Then these feature points are matched using normalized cross correlation. The RANSAC method [14] is used to eliminate the outliers and estimate the homographies. After spatial registration, exposure time is estimated using least-squares estimation. These estimates are then fine-tuned using [11], which is based on the Levenberg–Marquardt nonlinear optimization technique. Note that the feature-based registration approach may fail when the exposure time difference is significant, that is, when same features are not extracted as a result of saturation. Alternatively, it is possible to use mutual information for registration. Mutual information has been successfully used for intermodal registration [15]. Our experiments show that it is also effective for registering differently exposed images. The major drawback of mutual information is the computational complexity.

Fig. 4. Images of the same scene captured with a Canon G5 digital camera. These are three of the 15 images used in restoration. The highest exposure time is 1/50 s, and the lowest is 1/800 s.

There are various methods available in the literature to estimate camera response function [2], [8], [10]. In our experiments, we use the nonparametric camera response estimation method in [10].

C. Alternative Approaches

There are possible alternatives to the proposed approach. For example, one may apply super-resolution reconstruction to images with same exposure time and then fuse the reconstructed images to increase the dynamic range. This is definitely a suboptimal approach where the correlation among all images is not utilized. In [16], Capel and Zisserman adopted a linear photometric model, which uses global gain and offset terms for photometric changes. After normalizing the observations with the gain and offset factors, standard super-resolution reconstruction algorithms are applied. This is a good approach, unless there is saturation. We included results of this linear photometric model (LPM) algorithm in our experiments.

In case of saturation, it is possible to modify this model by estimating the saturated regions in the images and eliminating them during the restoration. This can be achieved by masking out the saturated regions. Saturated regions are chosen as the regions with the maximum pixel intensity. In this letter, we also included results of this modified linear photometric model (MLPM) algorithm.

D. Experiments

- In this experiment, we use all 28 images of the half-synthetic data set. We provide reconstruction results of three approaches. The first approach is the LPM algorithm, which does not take saturation into account. The second approach is the MLPM algorithm that uses a binary mask to eliminate the information coming from saturated regions. The third approach is the proposed algorithm. The resolution is increased by three. The noise variances σ_v^2 and σ_w^2 are empirically selected as 1 and 0.1, respectively. The point spread function is a 9×9 Gaussian window with standard deviation of 1.5. The number of iterations is 12. Fig. 5 shows three of the bilinearly interpolated input images and results of the three reconstruction approaches. The LPM algorithm loses information in saturated regions. The loss of contrast is obvious. The MLPM algorithm produces artifacts due to binary masking. Some contouring artifacts are also

a	b
c	d
e	f

Fig. 5. (a),(b),(c) Bilinearly interpolated observations with different exposure times. (d) Result of the LPM algorithm. (e) Result of the MLPM algorithm. (f) Result of proposed algorithm.

a	b
c	d

Fig. 6. (a) One of the interpolated observations. (b) Result of LPM algorithm. (c) Result of MLPM algorithm. (d) Result of proposed algorithm.

visible. On the other hand, the proposed algorithm has the best overall performance.

- This experiment is with the real data set. The resolution is increased by two. The point spread function is a 5×5 Gaussian window with standard deviation of 1.0. The rest of the restoration parameters are the same as in the previous experiment. Fig. 6 shows the results for this data set. For nonsaturated regions, all three algorithms performed similarly, as expected. However, the proposed algorithm outperforms the LPM and MLPM approaches considering the overall performance.

V. Conclusion

In this letter, we presented an approach to reconstruct high-spatial-resolution and high-dynamic-range images from multiple, possibly differently exposed images simultaneously. This is a generalization of super-resolution image reconstruction. We showed how to incorporate sensor nonlinearity into the reconstruction. Additive sensor noise and quantization noise are modeled; the way these noise terms come into picture is formulated. This also has implications in high-dynamic-range imaging. Experimental results with synthetic and real data sets prove the effectiveness of the approach. Some of the parameters used in the reconstruction were chosen heuristically. As a future work, we will investigate estimating optimal parameters and using other prior models.

References

[1] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 21–36, May 2003.

[2] P. E. Debevec and J. Malik, "Modeling and rendering architecture from photographs," in *Proc. ACM SIGGRAPH*, 1997, pp. 369–378.

[3] S. Mann, "Comparametric equations with practical applications in quantigraphic image processing," *IEEE Trans. Image Process.*, vol. 9, no. 8, pp. 1389–1406, Aug. 2000.

[4] 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.

[5] F. M. Candocia, "Joint registering images in domain and range by piecewise linear comparametric analysis," *IEEE Trans. Image Process.*, vol. 12, no. Apr., pp. 409–419, 2003.

[6] D. Capel and A. Zisserman, "Computer vision applied to super resolution," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 75–86, May 2003.

[7] B. K. Gunturk, Y. Altunbasak, and R. M. Mersereau, "Super-resolution reconstruction of compressed video using transform-domain statistics," *IEEE Trans. Image Process.*, vol. 13, no. 1, pp. 33–43, Jan. 2004.

[8] 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.

[9] X. Liu and A. E. Gamal, "Synthesis of high dynamic range motion blur free image from multiple captures," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 50, no. 4, pp. 530–539, Apr. 2003.

[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 Pattern Recognition*, vol. 1, 2001, pp. 842–849.

[11] F. M. Candocia, "Simultaneous homographic and comparametric alignment of multiple exposure adjusted pictures of the same scene," *IEEE Trans. Image Process.*, vol. 12, no. 12, pp. 1485–1494, Dec. 2003.

[12] A. Zomet, A. Rav-Acha, and S. Peleg, "Robust super-resolution," *Comput. Vis. Pattern Recognit.*, vol. 1, pp. 645–650, Dec. 2001.

[13] C. J. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. 4th Alvey Vision Conf.*, 1988, pp. 147–151.

[14] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Comm. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[15] P. Thevenaz and M. Unser, "Optimization of mutual information for multiresolution image registration," *IEEE Trans. Image Process.*, vol. 9, no. 12, pp. 2083–2099, Dec. 2000.

[16] D. Capel and A. Zisserman, "Super-resolution from multiple views using learnt image models," in *Proc. IEEE Int. Conf. Computer Vision Pattern Recognition*, vol. 2, Dec. 2001, pp. 627–634.