MAP Fusion Method for Superresolution of Images with Locally Varying Pixel Quality

Kio Kim, Nicola Neretti, Nathan Intrator

Department of Physics, Institute for Brain and Neural Systems, Brown University, Providence, RI 02912

Received 12 March 2007; accepted 7 April 2008

ABSTRACT: Superresolution is a procedure that produces a high-resolution image from a set of low-resolution images. Many of superresolution techniques are designed for optical cameras, which produce pixel values of well-defined uncertainty, while there are still various imaging modalities for which the uncertainty of the images is difficult to control. To construct a superresolution image from low-resolution images with varying uncertainty, one needs to keep track of the uncertainty values in addition to the pixel values. In this paper, we develop a probabilistic approach to superresolution to address the problem of varying uncertainty. As direct computation of the analytic solution for the superresolution problem is difficult, we suggest a novel algorithm for computing the approximate solution. As this algorithm is a noniterative method based on Kalman filter-like recursion relations, there is a potential for real-time implementation of the algorithm. To show the efficiency of our method, we apply this algorithm to a video sequence acquired by a forward looking sonar system. Int J Imaging Syst Technol, 18, 242–250, 2008; Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/ima.20137

Key words: superresolution; mosaicing; Kalman filter; fusion

I. INTRODUCTION

Superresolution has been an interesting problem for the development of digital imaging devices, and is currently an active area of research due to the widespread use of digital images in industrial, medical, and other everyday life applications. The term superresolution can vary depending on the context, but in general, superresolution means a procedure that produces a high-resolution image from a set of low-resolution images taken in slightly different conditions.

Producing a higher resolution image from a single low-resolution image is an underconstrained problem, and has been extensively studied as an interpolation problem (Crochiere and Rabiner, 1981; Schoenberg et al., 1994; Unser et al., 1995). When we have multiple low-resolution images, the number of constraints for the possible ground truth images is increased, and consequently, given a sufficient number of low-resolution images obtained from a target object, we can produce a high-resolution version of the target image that satisfies the constraints imposed by the low-resolution images.

The subcategories of superresolution algorithms include frequency domain methods (Tsai and Huang, 1984; Kim et al., 1990), nonuniform interpolation methods (Ur and Gross, 1992), projection onto convex sets (Stark and Oskoui, 1989; Tekalp et al., 1992), iterative back projection methods (Irani and Peleg, 1991, 1993; Mann and Picard, 1994), and Bayesian approaches (Cheeseman et al., 1994; Schultz and Stevenson, 1994; Shekarforoush et al., 1996). A detailed explanation of these algorithms can be found in the literature (Borman and Stevenson, 1998; Park et al., 2003; Farsiu et al., 2004). Many of the existing superresolution algorithms are developed for optical cameras that have regular receptive field size with homogeneous illumination intensity, such that all pixels in an image contribute to the superresolution with the same uncertainty. However, there exist many classes of images that have spatially varying uncertainty of pixel values and one needs new algorithms to achieve superresolution in such images. For example, changing illumination conditions can cause variability in the light intensity distribution of the images (Shashua, 1997), namely pixel values in poorly illuminated areas have lower signal-to-noise ratio (SNR), whereas in well-illuminated areas they have higher SNR (see Figs. 1d and 1e). In sonar images, such variation of illumination intensity is observed spatially as well as temporally (García et al., 2000). Although researchers have taken advantage of the varying illumination conditions in optical cameras to reconstruct 3D structure of a target object (Basri and Jacobs, 2001, 2003), the varying illumination condition can cause a problem when one wants to apply existing superresolution algorithms to the images under such circumstances (Chiang and Boult et al., 1997; Zhoa, 2004). Another example of images with varying uncertainty is the images of a forward looking sonar, a 2D sonar that produces images from a temporal series of 1D transducer responses. In these images, the spatial size of a pixel varies depending on the distance between the viewed area and the transducer of the sonar system (Kim et al., 2005). In that case, typical superresolution algorithms may fail to produce the desired results by assuming spatially uniform uncertainty of the pixel values (see Figs. 1b and 1c).

For the purpose of keeping track of uncertainty, a probabilistic approach is advantageous in that it can do the bookkeeping of the uncertainty of images. There are several probabilistic approaches to

Correspondence to: Kio Kim; e-mail: im@gmail.com

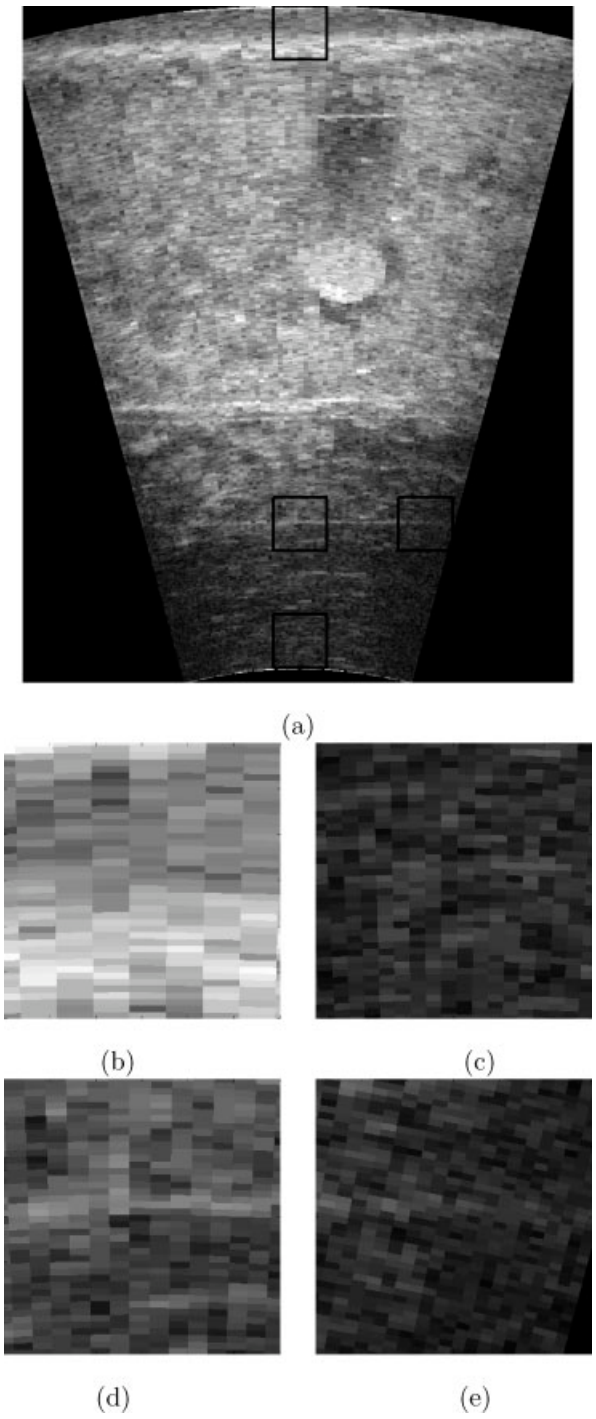


Figure 1. A DIDSON image showing inhomogeneous uncertainty. (a) A typical DIDSON frame. In a single frame of DIDSON sequence, some pixels are large (b) and some are small (c), and some areas are bright (d) and some are dark (e).

a superresolution problem including various maximum a posteriori (MAP) (Bose et al., 1994; Cheeseman et al., 1994; Schultz et al., 1996; Hong et al., 1997) and maximum likelihood algorithms (Tom et al., 1994; Tom and Katsaggelos, 1995). Among different MAP-based superresolution approaches, those with Gaussian Markov Random Field (GMRF) prior conditions are equivalent to Kalman filtering. There have been Kalman filter approaches toward solving superresolution problems (Dellaert et al., 1998; Elad and Feuer, 1999), which were particularly (Dellaert et al., 1998) focused on simplifying the equations via assuming the covariance matrix has negligible off-diagonal terms. Ready et al. (2006) implemented the same idea with an additional prior compensation algorithm to suppress the amplified artifacts due to neglecting the off-diagonal terms of the covariance matrix.

In this article, we derive an analytical expression for the MAP superresolution image with GMRF prior, and describe that this problem is equivalent to a Kalman filter. We solve the recursion relations of the Kalman filter equations with the same assumption as in the works of Dellaert et al. (1998) and Ready et al. (2006), namely very low correlation between nearby pixels, and derive expressions for the superresolution image and the covariance matrix as a function of all the low-resolution images. As this approach still retains the amplified artifacts which are resolved by regularization algorithms such as Tikhonov regularization (Schultz and Stevenson, 1994), we suppress the noise using the knowledge of the calculated estimation of the covariance.

The problem will be formulated in Section II. A MAP superresolution problem with GMRF prior is rewritten as a Kalman filtering problem in Section III, with the proposed solution of the recursion relations of Kalman filter equations in the Appendix. In Section IV we apply this method to produce a superresolution mosaic image from a video sequence produced by a forward looking sonar system called DIDSON (Dual-frequency Identification Sonar), which has both and irregular receptive field size and an inhomogeneous illumination condition. The result of this application is shown in Section V.

II. PROBLEM FORMULATION

A typical procedure for obtaining superresolution images includes image registration, image transformation, and image fusion. In the registration step, one obtains appropriate transformations between frames in the image sequence of interest. Depending on the image acquisition model, the transformation between two images is characterized by few parameters, which are in most of cases estimated by matching feature points in the two images. In doing this, outlier exclusion algorithms such as RANSAC (Random Sampling Consensus) (Fischler and Bolles, 1981) or LMedS (Least Median of Squares) (Rousseeuw, 1984) are employed to filter out wrong feature point matches. Once the transformation parameters between all the neighboring frames are found, one needs to combine those parameters appropriately to transform all the frames in the sequence to a single desired perspective. Then, in the image fusion step, all the observed and transformed image frames are fused into one image, which in general includes more information than the individual observed images. The problem we will address in this paper is the image fusion step, particularly when the images have spatially and temporally varying uncertainty.

We formulate the process of image acquisition as follows:

$$\beta_i = (\mathrm{L}_i \mathrm{W}_i \mathrm{M}_i)^\top \theta + \eta \equiv \mathcal{P}_i^\top \theta + \eta, \tag{1}$$

where M_i: the ith down-sampling operation; L_i: the ith illumination matrix; W_i: the ith point spread matrix; η: Gaussian noise.

The ith observed pixel value β_i is a scalar value of the intensity of a low-resolution pixel. For n frames of low-resolution images with m_{LR}-by-n_{LR} resolution, i is within the range of [1, $n \times m_{\mathrm{LR}} \times$

n_{LR}]. The vector θ is a column vector of size ($m_{HR} \times n_{HR}$)-by-1, which is a lexicographical representation of the ground truth image with the resolution of m_{HR}-by-n_{HR}. The matrices L_i and W_i are the illumination operator and the point spread operator, respectively, and both are ($m_{HR} \times n_{HR}$)-by-($m_{HR} \times n_{HR}$) diagonal matrices. In this article, the registration and transformation are assumed to be complete, so we have all the operating matrices such as L_i, W_i, and M_i known.

One thing to notice is that even if the variance of the Gaussian noise η is spatially homogeneous, the SNR in the observed image varies because L_i and W_i may vary spatially and temporally. Therefore, one needs to combine β_is so that the fusion image has the minimum amount of noise, or in other words, the maximum SNR.

We assume that θ is a random variable drawn from a multidimensional Gaussian distribution with the mean $\bar{\theta}$ and the covariance V

$$\mathcal{P}(\theta) \sim \mathcal{N}(\bar{\theta}, V), \tag{2}$$

where $p(\theta)$ denotes the probability of θ. Conditioned on more and more observations, V in (2) shrinks down, say, $\mathrm{Tr}(V_N) < \mathrm{Tr}(V_{N-1})$, when $\mathcal{P}(\theta|\beta_1 = P_1^\top\theta, \beta_2 = P_2^\top\theta, \ldots, \beta_N = P_N^\top\theta) \sim \mathcal{N}(\bar{\theta}_N, V_N)$, where $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix. Our goal is to estimate θ, the ground truth image, given a set of observed pixel values β_1, …, β_N, such that

$$\theta_{MAP} = \arg\max_{\theta} \mathcal{P}(\theta|\beta_1 = P_1^\top\theta, \beta_2 = P_2^\top\theta, \ldots, \beta_N = P_N^\top\theta). \tag{3}$$

III. METHODOLOGY

In this section, we formulate the MAP superresolution problem with GMRF prior in terms of a set of recursion relations, which is equivalent to a Kalman filtering. Using the assumption of negligible off-diagonal terms in the covariance matrix (Dellaert et al., 1998; Ready et al., 2006), we solve this recursion relations to obtain the expression of the superresolution image and its covariance matrix as functions of low-resolution images. Instead of suppressing the amplified artifacts using typical regularization methods, we achieve this via readjusting the image intensity so that the SNR is maximized through the image.

To obtain the distribution of θ conditioned on the ith observation, we rotate the coordinate system of θ so that

$$w_i \equiv P_i / ||P_i||_2, \tag{4}$$

is aligned to one of the axes in the coordinate system. By conditioning on the observed value, one can calculate the conditional distribution of θ with the variance along that axis shrunk in the rotated coordinate system, and then rotate the conditional distribution back to the original coordinate system. This procedure narrows down the variability of the estimated ground truth image. The detailed description of the procedure is provided later. For brevity, subscript i is ignored for R, w, $\bar{\theta}$, $\bar{y}$, $\bar{z}$, b, and V in this subsection.

A. The Conditional Probability.

In the rotated coordinate system, (2) is rewritten as

$$\mathcal{P}(R\theta) \sim \mathcal{N}(R^\top\bar{\theta}, RVR^\top). \tag{5}$$

The rotation matrix R is an arbitrary orthonormal matrix that includes w as one of the columns.

$$R = (W_\perp | w) \tag{6}$$

$$R^\top\theta = \begin{pmatrix} W_\perp^\top \\ w^\top \end{pmatrix}\theta \equiv \begin{pmatrix} y \\ z \end{pmatrix} \tag{7}$$

Note that b is the normalized version of the observation value β, i.e., $b_i \equiv \beta_i / \|P_i\|_2$. The conditional distributions of $y = W_\perp^\top\theta$ given the noiseless observation result $b = z = w^\top\theta$ is, as derived in the work of Press (1972),

$$\mathcal{P}(y|w^\top\theta = b) \sim \mathcal{N}(\bar{y} + V'_{12}V'^{-1}_{22}(b - \bar{z}), V'_{11,2}) \tag{8}$$

where

$$\begin{aligned} \bar{y} &= W_\perp^\top\bar{\theta} \\ \bar{z} &= w^\top\bar{\theta} \\ V'_{11} &= W_\perp^\top V W_\perp \\ V'_{12} &= W_\perp^\top V w \\ V'_{21} &= w^\top V W_\perp = V'^\top_{12} \\ V'_{22} &= w^\top V w \\ V'_{11,2} &= V'_{11} - V'_{12}V'^{-1}_{22}V'_{21}. \end{aligned}$$

However, in practice, the evaluation of z involves an observation error, say η_z, which we model as a Gaussian additive noise here;

$$b = w^\top\theta + \eta_z. \tag{9}$$

Then,

$$\begin{aligned} \mathcal{P}(z) &\sim \mathcal{N}(b, V_z) \\ \mathcal{P}(y|w^\top\theta = b - \eta_z) &\sim \mathcal{N}(\bar{y} + V'_{12}(V'_{22} + V_z)^{-1}(b - \bar{z}), V'_{11,3}), \end{aligned} \tag{10}$$

where V_z is the variance of the observation error η_z, and $V'_{11,3} \equiv V'_{11} - V'_{12}(V'_{22} + V'_z)V'_{21}$.

The conditional distribution of θ, which is also a Gaussian distribution, is obtained by rotating y back to the original coordinate system:

$$\mathcal{P}(\theta|b = w^\top\theta) = \mathcal{P}(R^\top R\theta|b = w^\top\theta).$$

This rotation is equivalent to inverting the operation of (7) after the conditioning;

$$\bar{\theta}^{\mathrm{opt}} = W_\perp\bar{y}^{\mathrm{opt}} + w\bar{z}^{\mathrm{opt}}, \tag{11}$$

where the superscript "opt" denotes the optimal estimate of the variable after the observation. In (11), $\bar{z}^{\mathrm{opt}}$ is the optimal fusion of the newly observed pixel value b and the previously known pixel value $w^\top\bar{\theta}$, of which the variances are V_z and V'_{22}, respectively;

$$\bar{z}^{\mathrm{opt}} = \frac{V'_{22}b + V_z w^\top\bar{\theta}}{V'_{22} + V_z}. \tag{12}$$

The optimal estimate of $\mathrm{w}_\perp \bar{\mathrm{y}}$ conditioned on the observation of b is

$$\begin{aligned}\mathrm{w}_\perp \bar{\mathrm{y}}^{\mathrm{opt}} &= \mathrm{w}_\perp \left(\mathrm{w}_\perp^\top \bar{\theta} + V'_{12}(V_{22} + V_z)^{\prime -1}(b - \bar{z}) \right) \\ &= \mathrm{w}_\perp \mathrm{w}_\perp^\top \left(\bar{\theta} + V\mathrm{w} \frac{b - \mathrm{w}^\top \bar{\theta}}{V'_{22} + V_z} \right) \\ &= (I - \mathrm{w}\mathrm{w}^\top) \left(\bar{\theta} - V\mathrm{w} \frac{(\mathrm{w}^\top \bar{\theta} - b)}{V'_{22} + V_z} \right) \\ &= \bar{\theta} - V\mathrm{w} \frac{(\mathrm{w}^\top \bar{\theta} - b)}{V'_{22} + V_z} - \bar{\mathrm{w}} z^{\mathrm{opt}} \end{aligned} \tag{13}$$

Thus, the expression for $\bar{\theta}^{\mathrm{opt}}$ is

$$\bar{\theta}^{\mathrm{opt}} = \bar{\theta} - V\mathrm{w} \frac{(\mathrm{w}^\top \bar{\theta} - b)}{V'_{22} + V_z}. \tag{14}$$

The optimal covariance matrix V^{opt} can be derived as described in the work of Press (1972);

$$V^{\mathrm{opt}} = \left(I - \frac{V\mathrm{w}\mathrm{w}^\top}{V'_{22} + V_z} \right) V. \tag{15}$$

Thus, we obtain a set of recursion formulae between the nth and the $(n + 1)$th mean values and covariance matrices as follows:

$$\bar{\theta}_{n+1} = \bar{\theta}_n - V_n \mathrm{w}_n \left(\frac{\mathrm{w}_n^\top \bar{\theta}_n - b_n}{\mathrm{w}_n^\top V_n \mathrm{w}_n + V_z} \right) \tag{16}$$

$$V_{n+1} = \left(I - \frac{V_n \mathrm{w}_n \mathrm{w}_n^\top}{\mathrm{w}_n^\top V_n \mathrm{w}_n + V_z} \right) V_n. \tag{17}$$

The recursion relations in (16) and (17) can be calculated given the initial estimation of the ground truth image θ_0 and the initial prior condition, or the covariance matrix V_0. As there is no prior knowledge on the ground truth image when no observation was performed, the initial ground truth image is a blank image with uniform pixel values. For the initial prior condition V_0, recent studies have suggested methods for choosing advanced priors (Pickup et al., 2004; Joshi et al., 2005; Jiji and Chaudhri, 2006). In this work, for the generality of application, it is assumed that the pixel values in θ are almost not correlated with its neighbors and there is no abrupt spatial change of variances. With this assumption, one can derive the following solution of $\bar{\theta}_{n+1}$ from the recursion relations in (16) and (17) as done in the Appendix:

$$\bar{\theta}_{n+1} \simeq \left(\sum_{i=1}^{n} \mathrm{L}_i^2 \mathrm{W}_i^2 + \nu_0 V_0^{-1} \right)^{-1} \left(\sum_{i=1}^{n} \mathrm{L}_i \mathrm{W}_i \mathrm{M}_i \beta_i \right). \tag{18}$$

This is the MAP solution of the ground truth image given, a GMRF prior with a covariance matrix V_0 with negligible off-diagonal terms.

Because of neglecting the off-diagonal terms of the covariance that provides information to suppress artifacts rising from noise, registration error, misshaped point spread function, etc., (18) is subject to a severe degradation. This problem is addressed by adjusting the intensity level so that the SNR becomes the maximum in all the pixels of the superresolution image: $\bar{\theta}_{n+1}^{\mathrm{MSNR}}$, the $(n + 1)$th fusion image with the maximum SNR (MSNR) is

$$\bar{\theta}_{n+1}^{\mathrm{MSNR}} \simeq \left(\sum_{i=1}^{n} \mathrm{L}_i \mathrm{W}_i + v_0 V_0^{-1} \right)^{-1} \left(\sum_{i=1}^{n} \mathrm{L}_i \mathrm{W}_i \mathrm{M}_i \beta_i \right). \tag{19}$$

The MSNR artifact suppression is extremely memory efficient in that it only requires two times of the size of the superresolution image, while it is resilient to noise or other registration errors because these errors are automatically penalized. In addition, because of the nature of the resulting equation in (19), the amount of required computation for adding a low-resolution image is constantly small regardless the number of all the observed images, which implies that any new observation can update the superresolution image in real time. These properties are desirable in many applications, for example, for on-board calculation of the superresolution mosaic image in a remotely operated vehicle or an autonomous underwater vehicle (AUV).

IV. APPLICATION

We applied (19) to DIDSON image sequence. In this section, DIDSON is briefly introduced, and the algorithm used to register DIDSON images is described.

A. Dual-Frequency Identification Sonar. DIDSON is a novel forward looking sonar system that is made up of a set of 1D acoustic lenses and an array of 96 transducers. The lens array is electronically controlled to focus the outgoing beam onto the targeted area, and the reflected beam onto the corresponding transducer. As this mechanism makes the beam forming of sonar complete without any further computation of signals, DIDSON can rapidly produce images at a frame rate as high as 20 frames per second. Because it operates with small power consumption and produces images at a high frame rate, the use of DIDSON is growing in surveillance, object tracking, investigation of underwater structures, and in many other areas.

A DIDSON image sequence is a good sample to test the proposed algorithm because the image has inhomogeneous uncertainty of pixel values. First, the pixel size varies within a single frame. Each transducer records a temporal series of wave amplitudes received from the designated angle, and the 96 transducers produce a image of 512-by-96 or 512-by-48 resolution for each frame (Belcher et al., 1999, 2001). The produced image is originally in the polar coordinate system, and is mapped to the Cartesian coordinate system for visualization. One pixel in the polar coordinate system has the same field-of-view and rangewise length, but when it is mapped to the Cartesian coordinate system, the actual pixel size and shape varies depending on the distance of the pixel from the camera (see Figs. 1b and 1c). Furthermore, the illumination profile varies in a frame for the following reason. The vertical viewing angle of DIDSON is limited to 10.8°. When the target object gradually leaves the vertical viewing angle, it receives less phonons and appears darker, eventually becoming completely dark when it is out of the viewing angle (see Figs. 1d and 1e). As a DIDSON system navigates underwater, the vertical displacement of an object fluctuates within the vertical viewing angle, which consequently causes the varying illumination profile.

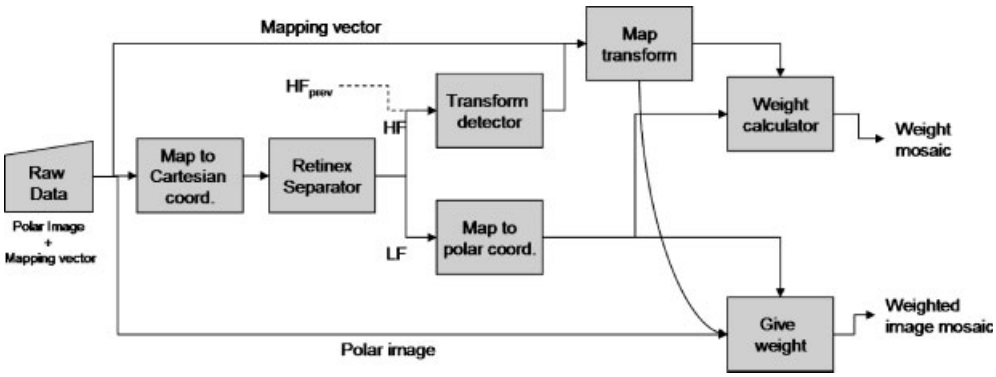


Figure 2. Block diagram for implementing the proposed superresolution algorithm to DIDSON images.

B. Procedure for the Registration and Mosaicing. In this subsection, we depict the full procedure for the registration and mosaicing steps of a DIDSON video sequence. Because the registration and the mosaicing are simultaneously done using the same variables and parameters, these steps are described as a whole.

The proposed algorithm for the registration and mosaicing of a DIDSON video sequence is as follows:

1. Determine the perspective of the desired mosaicing plane with respect to the perspective of the first frame.
2. Read the raw data (a lexicographical image vector and a mapping vector) for the first frame. Transform the mapping vector and point spread functions to the desired perspective in 0. Store the weighted image and the weight, where the weight is the point spread functions multiplied by the illumination profile.
3. Generate IM_1 by mapping the image vector to the Cartesian coordinates.
4. Separate the high frequency part HF_1 and the low frequency part LF_1 from IM_1.
5. Read the next raw data, and generate IM_2, HF_2, and LF_2 by repeating 2–3 for the next frame.
6. Find the transformation parameter T by registering HF_1 and HF_2.
7. From T, generate a mapping vector of IM_2 viewed from the predetermined perspective in 0.
8. From T, generate a map of the point spread functions $MPSF_2$ of the pixels in IM_2 viewed from the predetermined perspective.
9. Make a weight image for the current observed image by multiplying LF_2 and $MPSF_2$, and the weighted image by multiplying the IM_2 by the weight image. Add them to the existing weight image and weighted mosaic image produced in 1.
10. Rename IM_2, HF_2, and LF_2 with IM_1, HF_1, and LF_1, and repeat 4–8, until the end of the sequence.
11. Generate the superresolution mosaic image by point-to-point division of the sum of weighted images by the sum of weight.

The original data from DIDSON consists of a sequence of polar-coordinate images and the mapping information of the images from the polar coordinate system to the Cartesian coordinate system. The mapping vector is calculated based on the shape and the size of the viewed area. After the lexicographical image is mapped to the Cartesian coordinates, the mapped image IM_j is separated into two parts—the high frequency part HF_j and the low frequency part LF_j, where the subscript j denotes the jth frame. The lowpass filtered image LF_j is calculated by lowpass-filtering the mapped image, and is used as the estimated illumination intensity. The illumination process is modeled as a homomorphic filtering, and thus HF_j is calculated accordingly, say,

$$\mathrm{IM}_j(\mathrm{u}) = \mathrm{LF}_j(\mathrm{u}) \times \mathrm{HF}_j(\mathrm{u}) \tag{20}$$

$$\mathrm{HF}_j(\mathrm{u}) = \mathrm{IM}_j(\mathrm{u})/\mathrm{LF}_j(\mathrm{u}), \tag{21}$$

where IM_j (u) refers to the pixel value at u in the jth frame.

The registration of two images—namely, IM_j and IM_{j+1}—in the Cartesian coordinates is performed with HFs as follows:

1. Detect Kanade–Lucas–Tomasi corner points (Tomasi and Kanade, 1991) in HF_j.
2. For each corner point in HF_j, take a square patch around the point, and find the best matching point in HF_{j+1}.
3. Using the RANSAC algorithm (Fischler and Bolles, 1981), find the best parameters of the affine transformation between IM_j and IM_{j+1}.

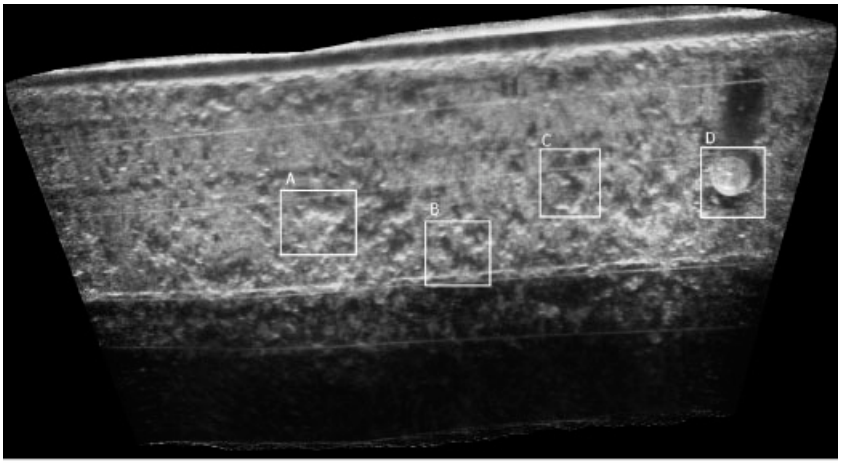
This procedure is based on the registration method described in detail in previous works (Kim et al., 2004, 2005).

The weight image is obtained by adding up the point-by-point multiplication of the PSF image and the illumination profile of each image. The point spread functions are assumed to be Gaussian, and the size is set to be proportional to the pixel size in the low-resolution images in the Cartesian coordinate system. The LF of the image is used for the illumination profile of an image.

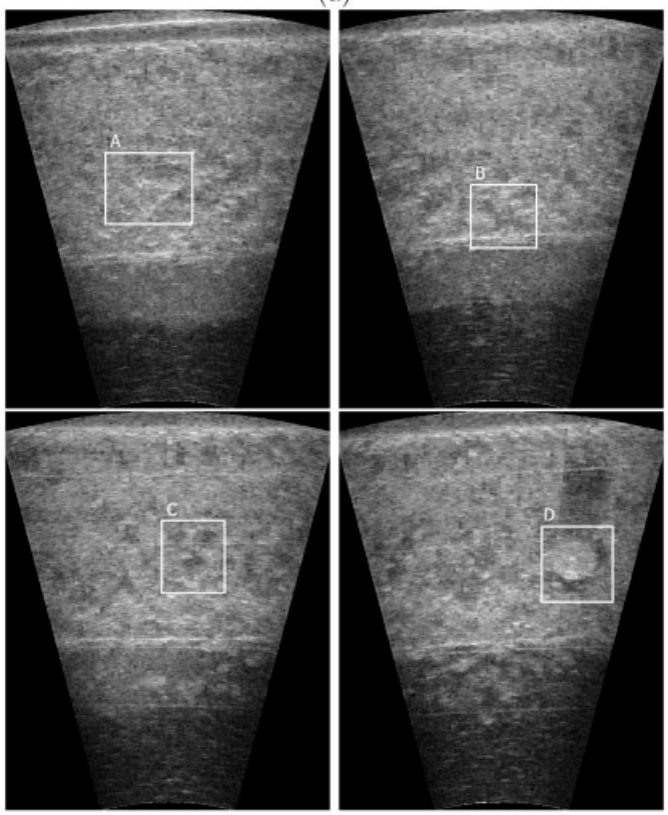
To construct the superresolution mosaic image in (19), we keep track of the sum of weights, and the sum of weighted images. After the sum of weights and the sum of weighted images are calculated for all of the frames in the sequence, the superresolution mosaic image is calculated by a point-to-point division of the sum of weighted images by the sum of weights. Figure 2 describes this fusion algorithm in a block diagram.

V. RESULTS

The algorithm has been tested on a DIDSON video sequence that was produced for ship hull inspection. An AUV equipped with a



(a)



(b)

Figure 3. Comparison of the mosaic image and the original images. (a) Mosaic of 80 frames of DIDSON images. (b) Common structures shown in the original frames are marked in original frames.

DIDSON system was put to move on a ship hull surface to scan the surface area while the vehicle wanders around.

The fusion image produced by the proposed algorithm is shown in Figure 3a. Figure 3a is the fusion image of 80 frames. Besides providing a wider view of the target object, one notices that the fusion of images improves the overall image quality in Figure 3, including the noise level, the intensity contrast, and the uniformity of illumination intensity.

Four structures that are commonly found in the mosaic image and in the original image are marked with rectangles. The structures in the original images are hard to recognize mainly due to the Rician noise (Wagner et al., 1983; Sijbers et al., 1998) and the consequent bias in the intensity profile (Kim et al., 2005). After the fusion, the image quality is drastically improved and the structures are strongly pronounced. For example, in the region around the marker D in Figures 3a and 3b, detailed features of the cylindrical object such as the reflection on the rim or border of the object can be much more easily identified in the improved image than the original image (see Fig. 4).

VI. CONCLUSION

In this article, we formulated a MAP based superresolution problem in terms of a conditional distribution with constraints imposed by

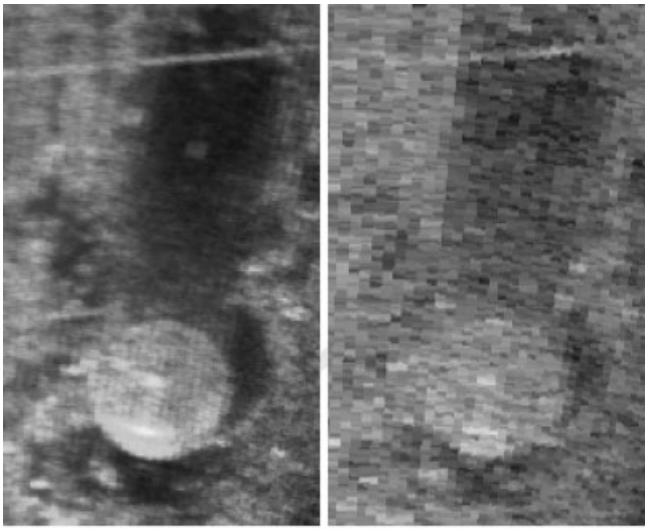


Figure 4. Comparison of the resolution enhanced image and the original image. A cylindrical object is viewed (a) in the superresolution image and (b) in the original image. In the superresolution image, about 10 frames of original DIDSON images are fused around a recognizable structure.

observed images. Instead of applying the Bayesian rule that involves a complicated integral or inversion of a large matrix, we sought an analytical solution to the problem with the minimum amount of approximation. In doing this, we employed a multivariate Gaussian model. Each time a new low-resolution image is acquired, the low-resolution image is fused into the high-resolution image, and this relationship is expressed as a set of recursion relations of the mean and the covariance matrix of the high-resolution image. Because of the dimensionality of the covariance matrix, additional approximations had to be made to yield a computationally feasible algorithm.

The idea of calculating the image of the MAP probability is equivalent to Kalman filtering, which provides the best estimation of the true reflectance values. In our work, a much simpler covariance matrix is assumed, which is basically simplifying the Kalman filtering, which consequently causes an amplification of noise in the resulting superresolution image. This has been noted by other researchers (Dellaert et al., 1998), and attacked with a locally iterative treatment for noise suppression (Ready et al., 2006). However, we noted that, in most of underwater imaging tasks, the users of the images are more interested in the overall layout of underwater circumstances rather than true reflectance profile. Therefore, we traded off accurate estimation of the true reflectance profile of underwater scenes for the best visualization of the region of interest.

One of the major novelties of this work is that it takes into account spatially and temporally varying illumination profiles and point spread functions when it produces the MAP high-resolution image from low-resolution images. In addition, this algorithm is free from iteration and it keeps track of the variance of the current high-resolution image based on recursion relations. These properties are desirable for a real-time implementation of the algorithm. This algorithm has been tested on a video sequence produced by a forward looking sonar system called DIDSON, and has shown satisfactory performance in generating a superresolution mosaic image from the sequence.

In summary, we proposed a new algorithm for a superresolution mosaicing using a conditional probability distribution, and implemented the algorithm to construct a superresolution mosaic image from a DIDSON video sequence. Significant noise reduction and resolution enhancement was observed during the process.

APPENDIX: SIMPLIFICATION OF THE RECURSION FORMULAE

Provided that the pixel values in the ground truth image θ are almost not correlated with its neighbors, and the variance of the pixel values vary smoothly in space, because w_n defined in (4) has its nonzero entries only around the nth observed pixel, it can be assumed that

$$V_n \mathrm{w}_n \simeq v_{y,n} \mathrm{w}_n, \tag{A1}$$

where $v_{y,n} \equiv \mathrm{w}_n^\top V_n \mathrm{w}_n$. With this assumption, we obtain the following from (17):

$$V_n - V_{n+1} = \frac{V_n \mathrm{w}_n \mathrm{w}_n^\top V_n}{v_{y,n} + v_{z,n}} = \frac{v_{y,n} \mathrm{w}_n \mathrm{w}_n^\top v_{y,n}}{v_{y,n} + v_{z,n}} = \frac{v_0 p_{y,n}^{-4} \mathrm{w}_n \mathrm{w}_n^\top}{p_{y,n}^{-2} + p_n^{-2}} \tag{A2}$$

$$= \frac{v_0 p_n^2 \mathrm{w}_n \mathrm{w}_n^\top}{p_{y,n+1}^2 p_{y,n}^2} \tag{A3}$$

$$= v_0 \mathrm{w}_n \mathrm{w}_n^\top \left(\frac{1}{p_{y,n}^2} - \frac{1}{p_{y,n+1}^2} \right), \tag{A4}$$

where p_n, $p_{y,n}$, and $p_{y,n+1}$ are defined as follows:

$$p_n^2 \equiv \frac{v_0}{v_{x,n}} \tag{A5}$$

$$p_{y,n}^2 \equiv \frac{v_0}{v_{y,n}} = \frac{v_0}{\mathrm{w}_n^\top V_n \mathrm{w}_n} \tag{A6}$$

$$p_{y,n+1}^2 \equiv p_{y,n}^2 + p_n^2. \tag{A7}$$

To reveal the relationship between V_{n+1} and $p_{y,n+1}$ we multiply both sides of (A4) by $\mathrm{w}_n^\top$ and w_n,

$$\mathrm{w}_n^\top V_n \mathrm{w}_n - \mathrm{w}_n^\top V_{n+1} \mathrm{w}_n = \frac{v_0}{p_{y,n}^2} - \frac{v_0}{p_{y,n+1}^2}, \tag{A8}$$

$$\therefore \mathrm{w}_n^\top V_{n+1} \mathrm{w}_n = \frac{v_0}{p_{y,n+1}^2} \equiv v_{y,n+1}. \tag{A9}$$

Plugging in (A6) and (A9) into (A3) and apply the assumption of slowly varying covariance matrix,

$$V_n - V_{n+1} \simeq v_0^{-1} V_n \mathrm{P}_n \mathrm{P}_n^\top V_{n+1}. \tag{A10}$$

Multiply V_{n+1}^{-1} and V_n^{-1} to both sides,

$$\begin{aligned} V_{n+1}^{-1} - V_n^{-1} &= v_0^{-1}\mathrm{P}_n\mathrm{P}_n^\top \\ \therefore V_{n+1}^{-1} &= V_n^{-1} + v_0^{-1}\mathrm{P}_n\mathrm{P}_n^\top \\ &= V_{n-1}^{-1} + v_0^{-1}\mathrm{P}_n\mathrm{P}_n^\top + v_0^{-1}\mathrm{P}_{n-1}\mathrm{P}_{n-1}^\top \\ &\cdots \\ &= V_0^{-1} + v_0^{-1}\sum_{i=1}^{n}\mathrm{P}_i\mathrm{P}_i^\top \end{aligned} \tag{A11}$$

$$\therefore V_{n+1} = v_0\left(\sum_{i=1}^{n}\mathrm{P}_i\mathrm{P}_i^\top + v_0V_0^{-1}\right)^{-1}, \tag{A12}$$

where V_0 is the initial covariance matrix, or the prior condition, which includes any GMRF priors.

In the same way, the assumption of slowly varying covariance matrix simplifies the recursion relation of the mean values in (16) as follows:

$$\begin{aligned} \bar{\theta}_{n+1} &= \bar{\theta}_n - \frac{V_n\mathrm{w}_n}{v_{y,n} + v_{z,n}}(\mathrm{w}_n^\top\bar{\theta}_n - b_n) \\ &= \bar{\theta}_n - \frac{p_n^2}{p_{y,n+1}^2}\mathrm{w}_n(\mathrm{w}_n^\top\bar{\theta}_n - b_n) \\ &= \bar{\theta}_n - v_0^{-1}V_{n+1}(\mathrm{P}_n\mathrm{P}_n^\top\bar{\theta}_n - p_n^2\mathrm{w}_nb_n) \\ &= \bar{\theta}_n - \left(\sum_{i=1}^{n}\mathrm{P}_i\mathrm{P}_i^\top + v_0V_0^{-1}\right)^{-1}(\mathrm{P}_i\mathrm{P}_i^\top\bar{\theta}_n - p_n^2\mathrm{w}_nb_n). \end{aligned} \tag{A13}$$

Multiplying both sides by $\sum_{i=1}^{n}\mathrm{P}_i\mathrm{P}_i^\top + v_0V_0^{-1}$,

$$\begin{aligned} \left(\sum_{i=1}^{n}\mathrm{P}_i\mathrm{P}_i^\top + v_0V_0^{-1}\right)\bar{\theta}_{n+1} &= \left(\sum_{i=1}^{n-1}\mathrm{P}_i\mathrm{P}_i^\top + v_0V_0^{-1}\right)\bar{\theta}_n + p_n^2\mathrm{w}_nb_n \\ &= \left(\sum_{i=1}^{n-2}\mathrm{P}_i\mathrm{P}_i^\top + v_0V_0^{-1}\right)\bar{\theta}_n \\ &\quad + p_n^2\mathrm{w}_nb_n + p_{n-1}^2\mathrm{w}_{n-1}b_{n-1} \\ &\cdots \\ &= \sum_{i=1}^{n}p_i^2\mathrm{w}_ib_i, \end{aligned} \tag{A14}$$

$$\therefore \bar{\theta}_{n+1} = \left(\sum_{i=1}^{n}\mathrm{P}_i\mathrm{P}_i^\top + v_0V_0^{-1}\right)^{-1}\left(\sum_{i=1}^{n}p_i^2\mathrm{w}_ib_i\right) \tag{A15}$$

$$= v_0^{-1}V_{n+1}\left(\sum_{i=1}^{n}p_i^2\mathrm{w}_ib_i\right). \tag{A16}$$

Remembering that $p_nb_n = \beta_n$ and $p_n\mathrm{w}_n = \mathrm{L}_n\mathrm{W}_n\mathrm{M}_n$, the recursion relation in (17) is reduced to

$$\bar{\theta}_{n+1} = \left(\sum_{i=1}^{n}\mathrm{L}_i^2\mathrm{W}_i^2 + v_0V_0^{-1}\right)^{-1}\left(\sum_{i=1}^{n}\mathrm{L}_i\mathrm{W}_i\mathrm{M}_i\beta_i\right). \tag{A17}$$

REFERENCES

R. Basri and D.W. Jacobs, Photometric stereo with general, unknown lighting, IEEE Conf Comput Vis Pattern Recognit 2 (2001), 374–381.

R. Basri and D.W. Jacobs, Lambertian reflectance and linear subspaces, IEEE Trans Pattern Anal Mach Intell 25 (2003), 218–233.

E.O. Belcher, H.Q. Dinh, D.C. Lynn, and T.J. Laughlin, Beamforming and imaging with acoustic lenses in small, high-frequency sonars, Proceedings of Oceans'99 MTS/IEEE, 1999, pp. 1495–1499.

E.O. Belcher, B. Matsuyama, and G.M. Trimble, Object identification with acoustic lenses, Proceedings of Oceans'01 MTS/IEEE, 2001, pp. 6–11.

S. Borman and R. Stevenson, Spatial resolution enhancement of low resolution image sequences—A comprehensive review with directions for future research, Technical Report, Laboratory for Image and Signal Analysis (LISA), University of Notre Dame, Notre Dame, IN, 1998.

N. Bose, H. Kim, and B. Zhou, Performance analysis of the TLS algorithm for image reconstruction from a sequence of undersampled noisy and blurred frames, Proc IEEE Int Conf Image Process 3 (1994), 571–575.

L.G. Brown, A survey of image registration techniques, ACM Comput Surv 24 (1992), 325–376.

P. Cheeseman, B. Kanefsky, R. Kraft, J. Stutz, and R. Hanson, Super-resolved surface reconstruction from multiple images, Fia-94-12, NASA Ames Research Center, Moffet Field, CA, 1994.

M.-C. Chiang and T.E. Boult, Local blur estimation and super-resolution, Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, June 1997, pp. 821–826.

R.E. Crochiere and L.R. Rabiner, Interpolation and decimation of digital signals-A tutorial review, Proc IEEE 69 (1981), 300–331.

F. Dellaert, C. Thorpe, and S. Thrun, Super-resolved texture tracking of planar surface patches, Proc IEEE/RSJ Int Conf Intelli Robotic Syst 1 (1998), 197–203.

M. Elad and A. Feuer, Super-resolution reconstruction of continuous image sequence, IEEE Trans Pattern Anal Mach Intell 21 (1999), 817–834.

S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, Advances and challenges in super-resolution, Int J Imaging Syst Technol 14 (2004), 47–57.

M.A. Fischler and R.C. Bolles, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, Commun ACM 24 (1981), 381–395.

R. García, X. Cufí, and L. Pacheco, Image mosaicking for estimating the motion of an underwater vehicle, Fifth IFAC Conference on Maneuvering and Control of Marine Craft, Aalborg, Denmark, 2000.

M. Hong, M. Kang, and A. Katsaggelos, An iterative weighted regularized algorithm for improving the resolution of video sequences, Proc IEEE Int Conf Image Process 2 (1997), 474–477.

M. Irani and S. Peleg, Improving resolution by image registration, CVGIP: Graph Models Image Process 54 (1991), 231–239.

M. Irani and S. Peleg, Motion analysis for image enhancement: Resolution, occlusion, and transparency, J Vis Commun Image Represent 4 (1993), 324–335.

C.V. Jiji and S. Chaudhuri, Single-frame image super-resolution through contourlet learning, EURASIP J Appl Signal Process 2006 (2006), 1–11.

M.V. Joshi, S. Chaudhuri, and R. Panuganti, A learning-based method for image super-resolution from zoomed observations, IEEE Trans Syst Man Cybern 35 (2005), 527–537.

S.P. Kim, N.K. Bose, and H.M. Valenzuela, Recursive reconstruction of high resolution image from noisy undersampled multiframes, IEEE Trans Acoust Speech Signal Process 38 (1990), 1013–1027.

K. Kim, N. Neretti, and N. Intrator, Acoustic camera image mosaicing and super-resolution, Proceedings of Oceans'04 MTS/IEEE, 2004, pp. 653–658.

K. Kim, N. Neretti, and N. Intrator, Mosaicing of acoustic camera images, IEE Proc Radar Sonar Navigat 152 (2005), 263–270.

S. Mann and R.W. Picard, Virtual bellows: Constructing high quality stills from video, Proceedings of International Conference on Image Processing, 1994, pp. 363–367.

S.C. Park, M.K. Park, and M.G. Kang, Super-resolution image reconstruction: A technical overview, IEEE Signal Process Mag 20 (2003), 21–36.

S.J. Press, Applied multivariate analysis, Holt, Rinehart and Winston Inc., San Francisco, CA, 1972.

L.C. Pickup, S.J. Roberts, and A. Zisserman, ''A sample texture prior for image super-resolution,'' In Proceedings of advances in neural information processing system 16, S. Thrun, L. Saul, and B. Schölkopf (Editors), MIT Press, Cambridge, MA, 2004, pp. 1587–1594.

B.B. Ready, C.N. Taylor, and R.W. Beard, A Kalman filter based method for creation of super-resolved mosaicks, Proceedings of IEEE International Conference on Robotics and Automation, Orlando, FL, 2006.

P.J. Rousseeuw, Least median of square regression, J Am Stat Assoc 79 (1984), 871–880.

I.J. Schoenberg, Cardinal interpolation and spline functions, J Approx Theory 2 (1969), 167–206.

R.R. Schultz and R.L. Stevenson, A Bayesian approach to image expansion for improved definition, IEEE Trans Image Process 3 (1994), 233–242.

R. R. Schultz and R.L. Stevenson, Extraction of high resolution frames from video sequences, IEEE Trans Image Process 5 (1996), 996–1011.

A. Shashua, On photometric issues in 3D visual recognition from a single 2D image, Int J Comput Vis 21 (1997), 99–122.

H. Shekarforoush, M. Berthod, J. Zerubia, and M. Werman, Subpixel Bayesian estimation of albedo and height, Int J Comput Vis 19 (1996), 289–300.

J. Sijbers, J den Dekker, P. Scheunders, and D.V. Dyck, Maximum-likelihood estimation of Rician distribution parameters, IEEE Trans Med Imaging 18 (1998), 357–361.

H. Stark and P. Oskoui, High resolution image recovery from imageplane arrays, using convex projections, J Opt Soc Am 6 (1989), 1715–1726.

A.M. Tekalp, M.K. Ozkan, and M.I. Sezan, High-resolution image reconstruction from lower-resolution image sequences and space varying image restoration, Proc IEEE Int Conf Acoust Speech Signal Process 3 (1992), 169–172.

B.C. Tom and A.K. Katsaggelos, Reconstruction of a high resolution image by simultaneous registration, restoration, and interpolation of low-resolution images. Proceedings of the International Conference on Image Processing, 1995, pp. 2539–2542.

B.C. Tom, A.K. Katsaggelos, and N.P. Galatsanos, Reconstruction of a high resolution image from registration and restoration of low resolution images. Proceedings of the International Conference on Image processing, 1994, pp. 553–557.

C. Tomasi and T. Kanade, Detection and tracking of point features, Technical Report CMU-CS-91-132, Carnegie Mellon University, Pittsburg, PA, 1991.

R.Y. Tsai and T.S. Huang, Multipleframe image restoration and registration, Adv Comput Vision Image Process 1 (1984), 317–339.

M. Unser, A. Aldroubi, and M. Eden, Enlargement or reduction of digital images with minimum loss of information, IEEE Trans Image. Process 4 (1995), 247–258.

H. Ur and D. Gross, Improved resolution from sub-pixel shifted pictures, CVGIP: Graph Model Image Process 54 (1992), 181–186.

R.F. Wagner, S.W. Smith, J.M. Sandrik, and H. Lopez, Statistics of speckle in ultrasound B-scans, IEEE Trans Sonics Ultrason 30 (1983), 156–163.

W.-Y. Zhao, Super-resolution with significant illumination change, Proceedings of the '04 International Conference on Image Processing (ICIP), 2004, pp. 1771–1774.