

MULTI-FRAME INFORMATION FUSION FOR GRAY-SCALE AND SPATIAL ENHANCEMENT OF IMAGES

Bahadır K. Gunturk, Yucel Altunbasak, and Russell M. Mersereau

Center for Signal and Image Processing
Georgia Institute of Technology
Atlanta, Georgia, 30332-0250

ABSTRACT

An imaging system produces a degraded measurement of a real-valued quantity that varies temporally, spectrally, and spatially. When there are multiple measurements of the same scene, it is possible to combine the non-redundant information in those measurements and produce an improved image that has more information than any of the measurements does alone. This type of reconstruction is known as image fusion. Depending on the diversity of information in the measurements, it is possible to achieve temporal, spectral, spatial, and gray-scale improvements with an image fusion algorithm. In this paper, we are proposing an image fusion algorithm that produces an image of higher spatial and gray-scale information from multiple measurements. The algorithm estimates the spatial and illumination correlation between multiple measurements, and employs a set-theoretic reconstruction technique.

1. INTRODUCTION

An image is a two-dimensional projection of a real-valued scene $q(t; \lambda; x, y, z)$ that is a function of time t , spectral wavelength λ , and space (x, y, z) . During the imaging process, the quantity $q(t; \lambda; x, y, z)$ is degraded in several ways. The degradation may occur on the domain $(t; \lambda; x, y, z)$ and/or on the range q . By combining multiple measurements, image fusion algorithms produce a composite (or a set of composites) that has more information than any of the individual measurements. With image fusion, we can improve resolution and/or extent in the four domains: temporal, spectral, spatial, and gray-scale.

A. Temporal Improvement: Frames of a video sequence are captured at a certain frame rate, and events that occur faster than the frame rate are not visible or else are observed incorrectly. It is possible to catch events that are faster than the frame rate when the motion can be modeled [1] and/or when there are multiple cameras that capture the scene with different time offsets or frame rates [2].

B. Spectral Improvement: Spectral filters are used to sample a certain portion of the spectrum. When the purpose is to produce a color picture, three different spectral filters usually provide sufficient visual quality for human viewing. However, in multi-spectral and hyper-spectral imagery, high spatial and spectral resolution is desirable [3].

C. Spatial Improvement: The result of an imaging process is two-dimensional data that is limited in its spatial extent and resolution. By combining multiple observations, it is possible to improve

the spatial information content. The reconstruction algorithm can seek three-dimensional or two-dimensional improvement. For instance, super-resolution reconstruction improves (2D) spatial resolution by sub-pixel registration of images [4, 5].

D. Gray-Scale Improvement: Image sensors produce continuously varying voltages that are proportional to the amount of light falling on them. Because of the limited dynamic range of the sensors, only a portion of the real dynamic range is captured. This is not the only source of information loss in the gray-scale domain, however. In order to process this data in the digital domain, the continuum of gray-scale values are quantized into a set of discrete values. The number of distinct gray levels is kept small for data storage and processing efficiency. As a result, a digital image is limited in both its gray-scale extent and its gray-scale resolution.

Recent work by Mann [6], Robertson *et al.* [7], and Candocia [8] demonstrate how to improve the dynamic range (gray-scale extent) of the image by combining images captured with different exposure times. Mann registered the images in range as a weighted sum of the pixel intensities. The weights were determined as a measure of reliability of the observed data. In Robertson [7] and Candocia [8] works, maximum likelihood and least squares estimation are used, respectively.

In this paper, we propose a set-theoretic image fusion algorithm that improves resolution and extent in both the gray-scale (range) and spatial domains. Although there has been a considerable number of super-resolution algorithms proposed, they do not consider improving the range (gray-scale extent and resolution). However, the observations may provide diverse information about range due to changes of illumination in the scene or imaging device adjustments such as exposure time, gain, or white balance, which can be used to produce composite images of higher dynamic range (gray-scale extent) and tonal fidelity (gray-scale resolution) in addition to higher spatial resolution and extent. The proposed algorithm can be considered as a generalization of super-resolution algorithms.

2. IMAGING MODEL

In this section, we use a more general image acquisition model (than the ones used in current super-resolution algorithms) that includes the factors that act in the gray-scale domain (such as the exposure time, white balance adjustment, saturation) in addition to the factors that act in the spatial domain (such as the point spread function of the sensors and sampling). The overall model is illustrated in Figure 1. Light coming from the spectrally and spatially varying scene $q(t; \lambda; x, y, z)$ is passed through an optical system to

This work was supported in part by the Office of Naval Research (ONR) under Award N00014-01-1-0619 and by the National Science Foundation under Award CCR-0113681.

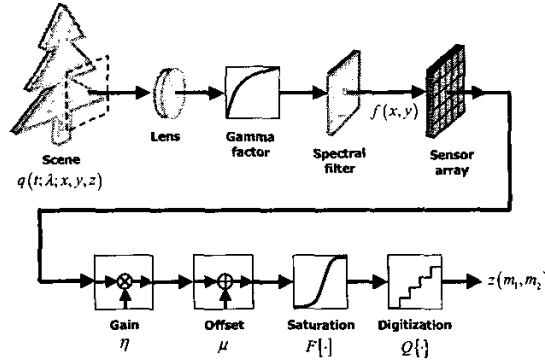


Fig. 1. Imaging model includes effects acting on the gray-scale domain as well as on the spatial domain.

form a two-dimensional real-valued function $f(x, y)$ on a sensor array. It is assumed that the scene is static (time-invariant) during the exposure time. A spectral filter captures a certain portion of the light spectrum. The sensors may have a nonlinear response to the amount of light falling on the sensor surface. This nonlinearity is usually by an exponential function, which is known as the gamma factor. Most commercial cameras have a built-in gamma correction circuit that linearizes the relationship between the impinging light intensity and the image output level. The rest of the imaging system converts the signal $f(x, y)$, which is continuous in the spatial and gray-scale domains, into a digital image. The process can be separated into two parts. The first part is a sensor array that converts the continuous signal into a discrete array using its discrete sensor elements. Each sensor element is an analog device with nonzero physical dimensions and produces a real-valued pixel intensity. The second part of the process acts on these pixel intensities. The camera may adjust the exposure time and white balance (automatically or manually by the user), which affect the pixel intensities as gain and offset terms, respectively. The pixel intensities are limited to a certain dynamic range and are quantized to a certain number of bits. The result is a digital image $z(m_1, m_2)$.

Current super-resolution algorithms try to reconstruct the signal $f(x, y)$ without modeling the processes acting on the gray-scale domain of the images. In practice, the signal is reconstructed on a discrete grid (n_1, n_2) instead of the continuous coordinates (x, y) , that has a finer resolution than the (m_1, m_2) grid. Incorporating the motion between observations, the super-resolution algorithms model the imaging process as a linear mapping between a high-resolution input signal $f(n_1, n_2)$ and low-resolution observations $z_i(m_1, m_2)$. This mapping includes motion (of the camera or the objects in the scene) and blur (caused by the point spread function of the sensor elements and the optical system); it can be formulated as

$$z_i(m_1, m_2) = \sum_{n_1, n_2} h_i(m_1, m_2; n_1, n_2) f(n_1, n_2), \quad (1)$$

where (n_1, n_2) and (m_1, m_2) are the discrete coordinates of the high- and low-resolution images, respectively, i is the observation number, and $h_i(m_1, m_2; n_1, n_2)$ is the linear mapping that incorporates motion, point spread function (PSF), and downsampling.

Details of such modeling can be found in [4, 5].

The model formulated in (1) assumes that the observations are captured under the same illumination conditions and camera settings. However, this assumption is not always valid. Including the effects acting on range (such as exposure time, gain, offset, etc.), we can extend the equation given in (1) as follows:

$$z_i(m_1, m_2) = Q \left\{ F \left[\eta_i \left(\sum_{n_1, n_2} h_i(m_1, m_2; n_1, n_2) f(n_1, n_2) \right) + \mu_i \right] \right\}, \quad (2)$$

where $F[\cdot]$ is the dynamic range saturation function, $Q\{\cdot\}$ is the gray-scale quantizer, and η_i and μ_i are the illumination gain and offset, respectively. In this model, the illumination gain and offset are assumed to be spatially uniform. This is a valid assumption unless there is non-uniform illumination change within the scene. For the case of non-uniform illumination effects, the parameters η_i and μ_i can be modeled as functions of the spatial coordinates.

3. JOINT GRAY-SCALE AND SPATIAL DOMAIN ENHANCEMENT

In this section we present a new image fusion algorithm for joint gray-scale and spatial enhancement. We define constraint sets using the quantization bounds of the pixel intensities, and employ a projections onto convex sets (POCS)-based algorithm to produce an image of higher spatial and gray-scale information content.

3.1. Constraint Sets from Amplitude Quantization

As formulated in (2), dynamic range compression and quantization introduces quantization error in the pixel intensities of the observations. Let $\delta_i(m_1, m_2)$ be the gray-scale quantization error at pixel (m_1, m_2) of the i th observation. Then we can rewrite (2) as

$$z_i(m_1, m_2) = \eta_i \left(\sum_{n_1, n_2} h_i(m_1, m_2; n_1, n_2) f(n_1, n_2) \right) + \mu_i + \delta_i(m_1, m_2). \quad (3)$$

Although the exact value of the quantization error cannot be determined from the measurements, the bounds within which it lies can be determined when the saturation curve of the camera is known. As illustrated in Figure 2, the input intensities are mapped by the nonlinear saturation function and then quantized to a certain number of bits, typically eight. For each intensity level, there is a lower bound $T_l(\cdot)$ and an upper bound $T_u(\cdot)$ within which the input intensity lies; and these bounds can be determined from the saturation curve.

We now define constraint sets using the bounds of the gray-scale data, and employ a POCS-based algorithm. In the POCS technique, each piece of information (arising from the observed data or prior knowledge) is associated with a constraint set in the solution space; the intersection of these sets represents the space of acceptable solutions. By projecting an initial estimate onto these constraint sets iteratively, a solution closer to the original signal is obtained. In this problem, the constraint sets arise from the quantization bounds of pixel intensities. Let $T_l(I)$ and $T_u(I)$ be the lower and upper quantization bounds for a measured pixel intensity I , and let $x(n_1, n_2)$ be an estimate of the high-quality image

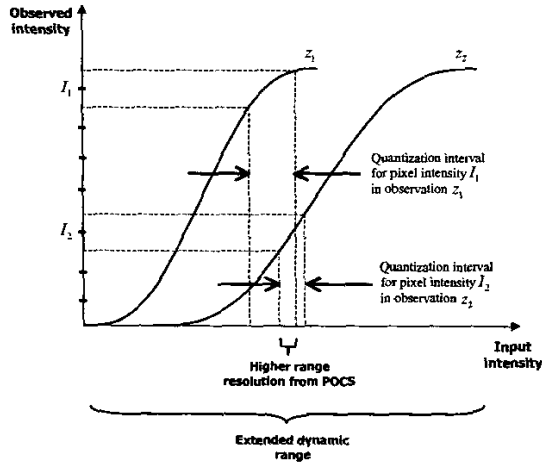


Fig. 2. For each observed intensity level, there is a lower bound and an upper bound within which the input intensity lies. It is possible to increase the gray-scale extent and resolution when there are multiple observations of different range spans.

$f(n_1, n_2)$. Then we can write the constraint set on any observed pixel $z_i(m_1, m_2)$ as follows:

$$C[z_i(m_1, m_2)] = \left\{ \begin{array}{l} f(n_1, n_2) : \\ T_l(z_i(m_1, m_2)) < \hat{z}_i(m_1, m_2) < T_u(z_i(m_1, m_2)) \end{array} \right\}, \quad (4)$$

where $\hat{z}_i(m_1, m_2)$ is the calculated pixel intensity derived from $x(n_1, n_2)$:

$$\hat{z}_i(m_1, m_2) = \eta_i \left(\sum_{n_1, n_2} h_i(m_1, m_2; n_1, n_2) f(n_1, n_2) \right) + \mu_i. \quad (5)$$

When there are multiple observations, the intersection of the constraint sets may result in finer range resolution and dynamic range. (See Figure 2.) In [4], a projections onto convex sets (POCS) algorithm for spatial-domain enhancement is presented. In that algorithm, bounds on the constraint sets are set to a fixed value, which is chosen heuristically, and the gray-scale effects are ignored. Here, we show that those bounds should actually be a function of observed pixel intensity, that they should be chosen using the camera response function. Moreover, the illumination changes and the camera settings should also be considered in the reconstruction.

3.2. Projection Operations

For each pixel in the observations, we can define a constraint set; and we project an initial estimate onto these constraint sets iteratively to update the initial estimate. The result is a solution of higher resolution in both the gray-scale and spatial domains.

The projection operator should update $x(n_1, n_2)$ in such a way that the constraint given in (4) is satisfied. We design the projection operator so that it projects the estimate $x(n_1, n_2)$ onto the bounds of the constraint sets. For instance, if $\hat{z}_i(m_1, m_2)$ is larger than the upper bound $T_u(z_i(m_1, m_2))$, then the projection operator

updates the estimate $x(n_1, n_2)$ such that the new $\hat{z}_i(m_1, m_2)$ (obtained from the updated $x(n_1, n_2)$) is equal to $T_u(z_i(m_1, m_2))$. After some simple algebra, the projection operator onto a constraint set $C[z_i(m_1, m_2)]$ can be found as

$$P_{C[z_i(m_1, m_2)]}[x(n_1, n_2)] = x(n_1, n_2) + \left\{ \begin{array}{l} \left(\frac{T_u(z_i(m)) - \hat{z}_i(m) - \mu_i}{\eta_i} \right) \bar{h}_i(m; n), \hat{z}_i(m) > T_u(z_i(m)) \\ 0, \quad T_l(z_i(m)) < \hat{z}_i(m) < T_u(z_i(m)) \\ \left(\frac{T_l(z_i(m)) - \hat{z}_i(m) - \mu_i}{\eta_i} \right) \bar{h}_i(m; n), \hat{z}_i(m) < T_l(z_i(m)) \end{array} \right\} \quad (6)$$

where $\mathbf{m} \equiv (m_1, m_2)$, $\mathbf{n} \equiv (n_1, n_2)$, $x(n_1, n_2)$ is an estimate of the original scene $f(n_1, n_2)$, $\hat{z}_i(m_1, m_2)$ is the calculated intensity from the estimate $x(n_1, n_2)$ as given in (5), and $\bar{h}_i(m; n)$ is the normalized blurring function:

$$\bar{h}_i(m; n) \equiv \frac{h_i(m_1, m_2; n_1, n_2)}{\sum_{n_1, n_2} |h_i(m_1, m_2; n_1, n_2)|^2}. \quad (7)$$

4. IMPLEMENTATION

In this section we summarize our implementation of the algorithm. Given a set of images, we estimate the motion and illumination parameters, register the images in spatial position range and range to obtain an initial estimate, and then project the initial estimate onto the constraint sets (defined for each pixel in the observations) iteratively. In our implementation we used the six-parameter affine motion model. To determine the affine parameters, we use the Harris corner detector [9] to select a set of points in the reference image. We then find the correspondence points in the second image using normalized cross correlation with quarter-pixel accuracy. A least-mean-square estimate of the affine parameters can be determined from the motion vectors determined at these points. Once the affine parameters are determined, the images are spatially registered and the illumination parameters η_i and μ_i are estimated, again using minimum least-mean-square estimation. The initial estimate is obtained by bilinearly interpolating one of the observations, and then extending the dynamic range by registering the other images in range. To perform range registrations, the images are first spatially registered (by warping on to the reference image using the affine parameters), and then registered in range by taking a weighted sum of the pixels. The weight function is chosen to be a Gaussian-like function; its mean is set to the middle point of the dynamic range, which is a result of the relative reliability of the midrange intensities [7].

5. EXPERIMENTAL RESULTS

We provide the results of an experiment to demonstrate the proposed method. We captured a video sequence using a Sony DCR-TRV20 digital camcorder. While capturing the sequence, we increased the exposure time manually. We used three images from the sequence in the reconstruction. These images are given in figures 3(a), 3(b), and 3(c). The image in Figure 3(a) was captured with a relatively short exposure time. In Figure 3(a) the low-contrast details in the lighter regions can be seen better, while in Figure 3(c) the tonal fidelity is higher for the darker regions.

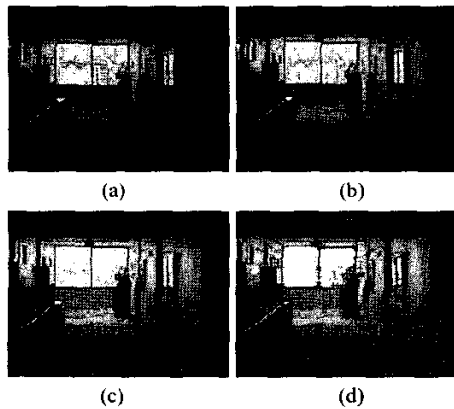


Fig. 3. Images from the first sequence. (a) First image. (b) Second image. (c) Third image. (d) Corners detected in the third image.

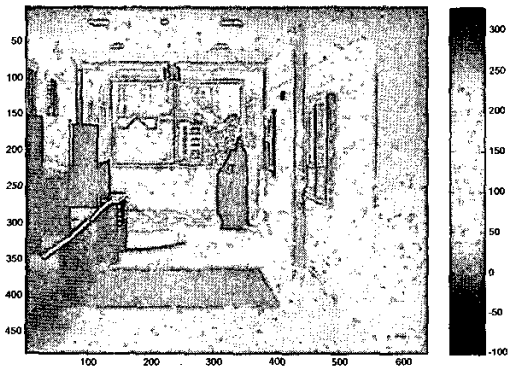


Fig. 4. Reconstructed image.

We applied the proposed algorithm to these images. The corners detected for the reference images are depicted in Figure 3(d). The correspondence points are found using normalized cross correlation with quarter-pixel accuracy, a block size of 8 pixels, and a search range of 10 pixels. The PSF is taken as a 7×7 Gaussian window with a standard deviation of one. Once the initial estimates are obtained, they are projected onto the constraint sets iteratively. The number of iterations is set to seven. The reconstructed image is given in Figure 4. It has higher dynamic ranges than any of the observations does by itself. (The pixel intensity range is given as a side bar.) The low-contrast regions also become more clear in the reconstructed images. After getting the high-dynamic-range image, certain portion of its range can be chosen for display purposes. In Figure 5, we show zoomed regions from these images. They are also scaled in intensity to range $[0 - 255]$. Close examination of these figures shows that the spatial resolution has also been improved during the reconstruction.

6. CONCLUSIONS

In this paper, we proposed an image fusion algorithm that improves resolution and extent in both the gray-scale and spatial do-

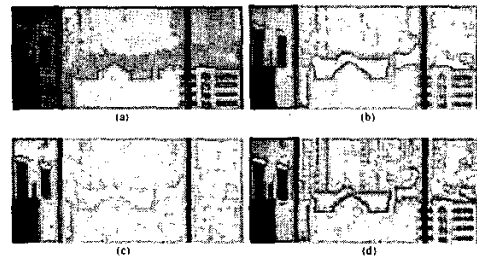


Fig. 5. Zoomed regions. (a) First image. (b) Second image. (c) Third image. (d) Reconstructed image scaled to intensity range $[0-255]$.

main. (Here we did not demonstrate increasing spatial extent, which is relatively straightforward.) We included the gray-scale effects of the image acquisition process and proposed a projections onto convex sets based reconstruction algorithm. The constraint sets are obtained from the gray-scale quantization information. Including the range effects, the algorithm can be considered as a generalization of spatial super-resolution algorithms.

7. REFERENCES

- [1] M. A. Robertson and R. L. Stevenson, "Temporal resolution enhancement in compressed video sequences," in *EURASIP Journal on Applied Signal Processing: Special Issue on Non-linear Signal Processing*, December 2001.
- [2] E. Shechtman, Y. Caspi, and M. Irani, "Increasing space-time resolution in video," in *European Conference on Computer Vision (ECCV)*, May 2002.
- [3] R. B. Gomez, A. Jazaeri, and M. Kafatos, "Wavelet-based hyperspectral and multispectral image fusion," in *2001 SPIE OE/Aerospace Sensing, Geo-Spatial Image and Data Exploitation II*, April 2001.
- [4] 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.
- [5] 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.
- [6] S. Mann, "Comparametric equations with practical applications in quantigraphic image processing," *IEEE Trans. Image Processing*, vol. 9, no. 8, pp. 1389–1406, August 2000.
- [7] M. A. Robertson, S. Borman, and R. L. Stevenson, "Dynamic range improvement through multiple exposures," in *Proc. IEEE Int. Conf. Image Processing*, 1999, vol. 3, pp. 159–163.
- [8] 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*, May 2002, vol. 4, pp. 3237–3240.
- [9] C. G. Harris and M. Stephens, "A combined corner and edge detector," in *Fourth Alvey Vision Conference*, 1988, pp. 147–151.