REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Service, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
31-05-2004	Final	01-May-2001 - 30-April-2004

4. TITLE AND SUBTITLE
"Super-Resolution Still and Video Reconstruction From MPEG Coded Video"

5a. CONTRACT NUMBER

5b. GRANT NUMBER
N00014-01-1-0619

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Altunbasak, Yucel

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Georgia Tech Research Corporation
Georgia Institute of Technology
Office of Sponsored Programs
Atlanta, GA 30332-0420

8. PERFORMING ORGANIZATION REPORT NUMBER
E-21-H52

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Ralph F. Wachter, ONR 311
Office of Naval Research
Ballston Centre Tower One
800 North Quincy Street
Arlington, VA 22217-5660

10. SPONSOR/MONITOR'S ACRONYM(S)
ONR

11. SPONSORING/MONITORING AGENCY REPORT NUMBER

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release; Distribution in Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
See attached page

20040804 073

15. SUBJECT TERMS
DCT Domain, Bayesian Superresolution, MPEG Video, Face Recognition, Image Reconstruction, Hyspectral Images

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
			N/A	87	Yucel Altunbasak
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPONE NUMBER *(Include area code)*
U					404-385-1341

Final Project Report

N00014-01-1-0619

by Yucel Altunbasak

Center for Signal and Image Processing
Georgia Institute of Technology
Atlanta, Georgia, 30332-0250
yucel@ece.gatech.edu

1 Summary of accomplishments

Our efforts in these twelve quarters resulted in four patents (two pending, two provisional applications), seven journal papers, seven conference papers and numerous professional and scholarly contributions, details of which are presented below.

- **Patents Pending**
  - Y. Altunbasak, B. Gunturk, and R. M. Mersereau, "Resolution enhancement and artifact reduction for MPEG video," filed with the Patent Office on April 26, 2002.
  - Y. Altunbasak and A. Patti, "A fast and robust model-based motion estimation strategy," filed with the Patent Office, Sept. 2002.
- **Provisional Patent Applications**
  - Y. Altunbasak and Y. C. Lee, "Multi-frame error concealment methods for transform-coded video," (GTRC ID 2509) filed with the USPTO on June 27, 2001
  - B. Gunturk, A. U. Batur, Y. Altunbasak, M. H. Hayes III, and R. M. Mersereau, Super-resolution for face recognition, (GTRC ID 2653) filed with the USPTO on May 3, 2002.
- **Journal Papers**
  - B. Gunturk, Y. Altunbasak, and R. M. Mersereau, "A DCT-domain bayesian resolution enhancement framework for MPEG video," IEEE Transactions on Image Processing, vol. 13, no. 1, pp. 33- 43, January 2004.
  - B. Gunturk, J. Glotzbach, Y. Altunbasak, R. W. Schafer, and R. M. Mersereau, "Demosaicking: Color filter array interpolation in single chip digital cameras," IEEE Signal Processing Magazine (Special Issue on Color Image Processing), September 2004.
  - B. Gunturk, Y. Altunbasak, and R. M. Mersereau, "Multiframe resolution-enhancement methods for compressed video," IEEE Signal Processing Letters, vol. 9, no. 6, pp. 170-174, June 2002.

- B. Gunturk, Y. Altunbasak, and R. M. Mersereau, "A multi-frame blocking artifact reduction method for transform coded video," IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, no. 4, pp. 269-276, April 2002.
- B. K. Gunturk, A. U. Batur, Y. Altunbasak, M. H. Hayes III, and R. M. Mersereau, "Eigenface-domain super-resolution for face recognition," IEEE Transactions on Image Processing, May 2003.
- Y. Altunbasak, A. J. Patti, and R. M. Mersereau, Super-resolution still and video reconstruction from MPEG coded video, IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, no. 4, pp. 217-227, April 2002.
- B. Gunturk, Y. Altunbasak, and R. M. Mersereau, Color plane interpolation using alternating projections, IEEE Transactions on Image Processing, vol. 11, no. 9, pp. 997-1013, September 2002.

- **Conference Papers**

  - 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, Barcelona, Spain, Sep. 2003.
  - B. Gunturk, A. Batur, Y. Altunbasak, M. H. Hayes III, and R. M. Mersereau, "Eigenface-based super-resolution for face recognition," IEEE Int. Conf. on Image Processing, vol. 2, pp. 845-848, Rochester, NY, September 2002.
  - B. Gunturk, Y. Altunbasak, and R. M. Mersereau, "A multi-frame blocking artifact reduction method for transform coded video," IEEE Int. Conf. Acoust. Speech Sign. Proc., vol. 3, pp. 1789-1792, Salt Lake City, UT, May 2001.
  - B. Gunturk, Y. Altunbasak, and R. M. Mersereau, "A bayesian resolution enhancement framework for transform coded video," Proc. IEEE Int. Conf. Image Proc., vol. 2, pp. 41-44, Thessaloniki, Greece, Oct. 2001.
  - B. Gunturk, Y. Altunbasak, and R. M. Mersereau, "Gray-scale resolution enhancement," IEEE Workshop on Multimedia Signal Processing, pp. 155-160, Cannes, France, Oct. 2001.
  - B. Gunturk, Y. Altunbasak, and R. M. Mersereau, Color plane interpolation using alternating projections, IEEE Int. Conf. on Acoustics Speech and Signal Processing, vol. 4, pp. 3333-3336, Orlando, FL, May 2002.
  - Toygar Akgun, Yucel Altunbasak and R. M. Mersereau, "Superresolution reconstruction of hyperspectral images," in Proc. IEEE Int. Conf. on Acoustics Speech and Signal Processing (ICASSP), Montreal, Canada, May 2004

- **Main Accomplishments**

  - Multi-frame Blocking Artifact Reduction Method for Transform Coded Video and its Relation to Superresolution (Section 4.1)
  - DCT-Domain Bayesian Superresolution Reconstruction for MPEG Video (Section 4.2)
  - Fast Motion Estimation Using Low-bit-resolution Images (Section 4.3)
  - Multi-Frame Gray-Scale Resolution Enhancement (Section 4.4)
  - Face Recognition from Video (Section 4.5)

- Effects of Camera Response Function and Illumination Changes in Multi-Frame Image Reconstruction (Section 4.6)
- Multi-Frame Information Fusion for Gray-Scale and Spatial Enhancement of Images (Section 4.7)
- Super-Resolution Reconstruction of Hyper-Spectral Images (Section 4.8)

2 Professional and Scholarly Contributions

- **Leadership Activities**
  - Technical program chair, IEEE Int. Conf. on Image Processing (ICIP'06), Atlanta, GA, 2006.
  - Advanced Signal Processing for Communications" Symposia co-chair, IEEE International Conference on Communications (ICC'03), May 11-15, 2003
  - Multimedia Networking Technical Track Chair, IEEE International Conference on Multimedia and Expo (ICME'04), June 27-30, 2004
  - Multimedia Networking Technical Track Chair, IEEE International Conference on Multimedia and Expo (ICME'03), July 6-9, 2003
  - Panel sessions chair, International Conference on Information Technology: Research and Education (ITRE'03), August 10-13, 2003
  - Session chair, Video Transcoding, IEEE Int. Conf. on Image Processing (ICIP'03), September 15, 2003
  - Session chair, Wireless Multimedia I, IEEE Int. Conf. Multimedia and Expo (ICME), Baltimore, MD, July 7, 2003
  - Session chair, Multimedia I, IEEE Int. Conf. on Communications (ICC'03), May 12, 2003
  - Session chair, Wireless networking, IEEE Int. Conf. on Communications (ICC'03), May 14, 2003
  - Session chair, Error Concealment, IEEE Int. Conf. on Image Processing (ICIP'02), Sept. 22-25, 2002
  - Session chair, Image/Video Coding, IEEE Int. Conf. Acoust. Speech Sign. Proc. (ICASSP'02), May 10-14, 2002
  - Session chair, Digital Video Processing, 32nd Asilomar Conference on Signals, Systems, and Computers, Nov. 1-4, 1998
- **Editorial Activities**
  - Associate Editor, IEEE Transactions on Image Processing, 5/26/2002-6/1/2005
  - Associate Editor, IEEE Transactions on Signal Processing, 7/31/2003-7/31/2005
  - Area Editor, Signal Processing: Image Communications, 1/1/2001-1/1/2004
  - Associate Editor, Circuits, Systems, and Signal Processing, 6/21/2000-12/31/2002

- Guest Editor, EURASIP Image Communications Special Issue on "Recent Advances in Wireless Video"
- Panelist and proposal reviewer for NSF, June 2001, Feb. 2003, and May 2003
- Panelist and proposal reviewer for ARO, March 2001

- **Membership Activities**
  - Advisory Board, Department of Electrical Engineering, Bilkent University, Ankara, Turkey
  - Vice-president - North America, IEEE Communications Society Multimedia Communications Technical Society, 1/2004-1/2006
  - Elected member, IEEE Signal Processing Society Image and Multi-dimensional Signal Processing (IMDSP) Technical Committee, 5/2002-5/2008
  - Senior Member, IEEE, 2002
  - Technical program committee member, ICME'04, June 2004, Taipei, Taiwan
  - Technical program committee member, ITRE'04, June 2004, London, England
  - Technical program committee member, ICASSP'04, Montreal, Canada, May 2004
  - Technical program committee member, ICC'04, Paris, France, June 2004
  - Technical program committee member, ICIP'03, Barcelona, Spain, September 2003
  - Technical program committee member, ICME'03, July 2003, Baltimore, MD
  - Technical program committee member, ICASSP'03, Hong Kong, China, April 2003
  - Technical program committee member, Wireless Personal Multimedia Communications (WPMC'02)
  - Technical program committee member, IEEE Int. Conf. on Image Processing (ICIP'02), Rochester, NY, September, 2002
  - Technical program committee member, ICIP'01, Thessaloniki, Greece, October 2001
  - Technical program committee member, IEEE International Symposium on Circuits and Systems (ISCAS'02), Scottsdale, Arizona, May 2002
  - Technical program committee member, ICASSP'01, Salt Lake City, Utah, May 2001

3 Honors and Awards

- Received NSF CAREER award (2002)
- Received "Outstanding Junior Faculty Award" at the School of Electrical and Computer Engineering, Georgia Tech (2003)
- Co-author for a conference paper that received the "best student paper award" at ICIP 2003

4 Main Accomplishments and Results

4.1 Multi-frame Blocking Artifact Reduction Method for Transform Coded Video

Transform coding is a popular and effective compression method for both still images and video sequences, as is evident from its widespread use in international media coding standards such as MPEG, H.263 and JPEG. The motion-compensated image (or the image itself) is divided into blocks and each block is independently transformed by a 2-D orthogonal transform to achieve energy compaction. The most commonly used transform is the discrete cosine transform (DCT). After the block transform, the transform coefficients undergo a quantization step. At low bit-rates, the DCT coefficients are coarsely quantized. This coarse quantization with independent quantization of neighboring blocks gives rise to blocking artifacts—visible block boundaries.

Several blocking artifact reduction methods have been proposed in the literature. Spatial filtering, iterative reconstruction techniques, and stochastic reconstruction techniques are among the blocking artifact reduction methods that have been proposed for still images. Although temporal information adds another dimension for video sequences, none of the prior art used this information effectively in blocking artifact reduction in video.

In this work, we develop a multi-frame restoration-based method that makes use of the spatial correlations between successive frames effectively. The proposed method constructs convex constraint sets *at each frame within a neighborhood of the frame of interest, using the motion between the frames and the quantization information extracted from the bit stream.* The method is based on the fact that, although the exact value of the quantization noise added to each DCT coefficient is not known, the range within which it lies can be determined from the bit stream. Incorporating the motion between the frames, we can define constraint sets not only at the current frame, but also at each frame within a small neighborhood of the current frame. By projecting the initial "blocky" frame onto these constraint sets successively, we can reconstruct a better estimate of the "original" frame—the one before the quantization step.

In order to derive the constraint sets, we first need to establish the relation between neighboring frames. The key is to use the intensity conservation assumption along the motion trajectories. Let $f(\mathbf{x}, t)$ denote the intensity of the continuous spatio-temporal video signal at spatial coordinate $\mathbf{x} \equiv [x_1, x_2]$ at time t. Pixel intensities of any two video frames can be related to each other through the motion vectors. Denoting $\mathbf{M} \equiv [M_1(\mathbf{x}, t_k; t_j), M_2(\mathbf{x}, t_k; t_j)]$ as the motion mapping between the frames at times t_k and t_j, we can write:

$$f(\mathbf{x}, t_k) = f(\mathbf{x} - \mathbf{M}, t_j). \tag{1}$$

We now proceed by relating the spatially continuous video frame at time t_j to the corresponding discrete frame. Letting $f(\mathbf{n}, j)$ denote the intensity of the jth discrete frame at the integer coordinate

$\mathbf{n} \equiv [n_1, n_2]$, we can write the spatially-continuous reconstruction as:

$$\begin{aligned} f(\mathbf{x}, t_j) &= \left[\sum_{\mathbf{n}} f(\mathbf{n}, j)\delta(\mathbf{x} - \mathbf{n})\right] * h_r(\mathbf{x}) \\ &= \int \sum_{\mathbf{n}} f(\mathbf{n}, j)\delta(\xi - \mathbf{n}) h_r(\mathbf{x} - \xi) d\xi \\ &= \sum_{\mathbf{n}} h_r(\mathbf{x} - \mathbf{n}) f(\mathbf{n}, j), \end{aligned} \tag{2}$$

where $h_r(\mathbf{x})$ is the reconstruction filter. Substituting Equation (2) into Equation (1), and discretizing k^{th} frame, we get:

$$f(\mathbf{m}, k) = \sum_{\mathbf{n}} h_r(\mathbf{m} - \mathbf{M} - \mathbf{n}) f(\mathbf{n}, j). \tag{3}$$

Now we model the operations that take place in the process of MPEG compression (*i.e.,* motion compensation, block-DCT calculation, and quantization). Motion compensation is simply the subtraction of an offset value $\hat{f}(\mathbf{m}, k)$ (predicted image) from $f(\mathbf{m}, k)$. The residual image is then passed through a series of 8×8 block-DCTs to result in DCT coefficients. This is followed by a quantization step which can be modeled as an addition of quantization error. Using Equation 3, the overall relation can be written as:

$$d_q(\mathbf{m}, k) = \sum_{\mathbf{n}} h_K(\mathbf{m}, k; \mathbf{n}, j) f(\mathbf{n}, j) - \hat{F}(\mathbf{m}, k) + Q(\mathbf{m}, k), \tag{4}$$

where $d_q(\mathbf{m}, k)$ is the quantized DCT coefficients, $h_K(\mathbf{m}, k; \mathbf{n}, j)$ is the block-DCT of $h_r(\mathbf{m} - \mathbf{M} - \mathbf{n})$, $\hat{F}(\mathbf{m}, k)$ is the block-DCT of $\hat{f}(\mathbf{m}, k)$, and $Q(\mathbf{m}, k)$ is the quantization error.

Although the exact value of $Q(\mathbf{m}, k)$ is not known, the range within which the DCT coefficient lies can be extracted from the MPEG bit stream. Based on this fact we define constraint sets $C(\mathbf{m}, k)$ on frame $f(\mathbf{m}, j)$. Defining $b_l(\mathbf{m}, k)$ and $b_u(\mathbf{m}, k)$ as the lower and upper bounds of the DCT coefficient at spatio-temporal location $(\mathbf{m}, k)$, $C(\mathbf{m}, k)$ can be written as:

$$C(\mathbf{m}, k) = \left\{ f(\mathbf{m}, j) : \left[\sum_{\mathbf{n}} h_K(\mathbf{m}, k; \mathbf{n}, j) f(\mathbf{n}, j) - \hat{F}(\mathbf{m}, k)\right] \in [b_l(\mathbf{m}, k), b_u(\mathbf{m}, k)] \right\}. \tag{5}$$

This equation shows how to define constraint sets on any frame j using the quantization information from another frame k. By projecting the "blocky" frame onto these constraint sets, the blocking artifacts can be reduced significantly.

This constraint set is defined simply by using the quantization bounds $[b_l(\mathbf{m}), b_u(\mathbf{m})]$ of one of the DCT coefficients as explained in the previous section. Projection of an image onto this constraint set amounts to simply finding the closest point on $C(\mathbf{m}, k)$.

Although this projection is likely to reduce the blocking artifacts, it does not guarantee a significant improvement since the "original" blocking-artifact-free image could be anywhere in the shaded region. Defining another constraint set could improve the quality significantly. As depicted in Figure 43, the second constraint set defined with help of the neighboring frame $(k + 1)$ reduces the region where the "original" image lies. Projecting the initial frame onto these convex sets successively produces a better result. By using additional frames we can impose more constraint sets onto the reconstructed frame and reduce the blocking artifacts further.

4.2 DCT-Domain Bayesian Superresolution Reconstruction for MPEG Video

With a Bayesian estimator, not only the source statistics, but also various regularizing constraints can be incorporated into the solution. Bayesian estimators have been frequently used for superresolution. However, in these approaches either the video source is assumed to be available in uncompressed form, or it is simply decompressed without considering the quantization process. Additive noise is considered as the only source of error. On the other hand, the projection onto convex sets (POCS) techniques treat the quantization error as the only source of error without considering the additive noise. Clearly, neither of these approaches provides a complete framework for superresolution. The model that we have developed combines the quantization process and the additive noise under a Bayesian framework.

We begin by analyzing the block diagram in Figure 1 depicting the video acquisition process. According to this model, a spatially and temporally continuous high-resolution input signal $f(\mathbf{x}, t)$ is affected by sensor and optical blurs. Sensor blur is caused by integrating over the finite nonzero sensor cell area, while optical blur is caused by the lens system. The blurred video signal is also integrated over time to capture nonzero time-aperture effects. After sampling on a low-resolution grid, discrete low-resolution frames $g_d(\mathbf{l}, k)$ are obtained.

We now add the MPEG compression stages to this model. As shown in Figure 2, the LR frame $g_d(\mathbf{l}, k)$ is motion compensated (*i.e.*, the prediction frame is computed and subtracted from the original to get the residual image), and the residual is transformed using a series of 8×8 block-DCTs to produce the DCT coefficients $d(\mathbf{m}, k)$. The DCT coefficients $d(\mathbf{m}, k)$ are then quantized to produce the quantized DCT coefficients $\tilde{d}(\mathbf{m}, k)$.

In the maximum *a posteriori* probability (MAP) formulation, the quantized DCT coefficients, $\tilde{d}(\mathbf{m}, k)$, the original high-resolution frame, $f(\mathbf{n}, t_r)$, and the block DCT of the additive noise are all assumed to be random processes. Denoting $p_{f(\underline{\mathbf{n}},t_r)|\tilde{d}(\underline{\mathbf{m}},k_1),\cdots,\tilde{d}(\underline{\mathbf{m}},k_p)}(\cdot)$ as the conditional probability density function (PDF) with $k_1, \cdots, k_p$ being the frames used in the reconstruction, the MAP estimate $\hat{f}(\mathbf{n}, t_r)$ is given by:

$$\hat{f}(\mathbf{n}, t_r) = \underset{f(\underline{\mathbf{n}},t_r)}{\arg\max} \left\{ p_{f(\underline{\mathbf{n}},t_r)|\tilde{d}(\underline{\mathbf{m}},k_1),\cdots,\tilde{d}(\underline{\mathbf{m}},k_p)}(\cdot) \right\}, \qquad \forall\, \mathbf{n}. \tag{6}$$

We used the <u>underline</u> notation (in $\mathbf{n}$ and $\mathbf{m}$) to emphasize that this PDF is the joint PDF, not the PDF at a specific location.

Using Bayes' rule, Equation 6 can be rewritten as:

$$\hat{f}(\mathbf{n}, t_r) = \underset{f(\underline{\mathbf{n}},t_r)}{\arg\max} \left\{ p_{\tilde{d}(\underline{\mathbf{m}},k_1),\cdots,\tilde{d}(\underline{\mathbf{m}},k_p)|f(\underline{\mathbf{n}},t_r)}(\cdot) p_{f(\underline{\mathbf{n}},t_r)}(\cdot) \right\}, \qquad \forall\, \mathbf{n}, \tag{7}$$

where we used the fact that $p_{\tilde{d}(\underline{\mathbf{m}},k_1),\cdots,\tilde{d}(\underline{\mathbf{m}},k_p)}(\cdot)$ is not a function of $f(\underline{\mathbf{n}}, t_r)$. Clearly we need to model the conditional PDF $p_{\tilde{d}(\underline{\mathbf{m}},k_1),\cdots,\tilde{d}(\underline{\mathbf{m}},k_p)|f(\underline{\mathbf{n}},t_r)}(\cdot)$ and the prior PDF $p_{f(\underline{\mathbf{n}},t_r)}(\cdot)$ in order to find the MAP estimate $\hat{f}(\mathbf{n}, t_r)$. If the additive noise $v(\mathbf{m}, k)$ is assumed to be an independent, identically distributed (IID) Gaussian process, it is possible to derive the conditional PDF $p_{\tilde{d}(\underline{\mathbf{m}},k_1),\cdots,\tilde{d}(\underline{\mathbf{m}},k_p)|f(\underline{\mathbf{n}},t)}(\cdot)$, analytically, which can be used with any prior image model. The resulting estimation problem can be implemented using an iterated conditional modes (ICM) scheme. (The details of the derivation and the implementation are presented in the attached paper.)

To demonstrate the efficacy of the proposed method we performed a set of experiments. In these experiments, the test images AERIAL (Figure 3) and WATCH (Figure 7) were jittered, blurred, downsampled, and corrupted by additive noise to produce multiple low-resolution frames. For blurring, the images were convolved with a 5×5 Gaussian kernel with a standard deviation of 1.5 pixels. The additive noise had a Gaussian distribution with a standard deviation 1.4. The aperture time was taken to be zero, and the lens point-spread function was assumed to be space invariant. The low-resolution frames were then compressed using an MPEG encoder operating at 1.5Mbits/sec to produce the low resolution video sequence. Figure 4 and Figure 8 show one of those low resolution frames from each of the sequences.

The first decoded frame for each video sequence was bilinearly interpolated. These images are depicted in Figures 5 and 9. We then applied our compressed-domain resolution enhancement method to upsample the first frames for both sequences. Figures 6 and 10 illustrate the results obtained by the proposed algorithm.

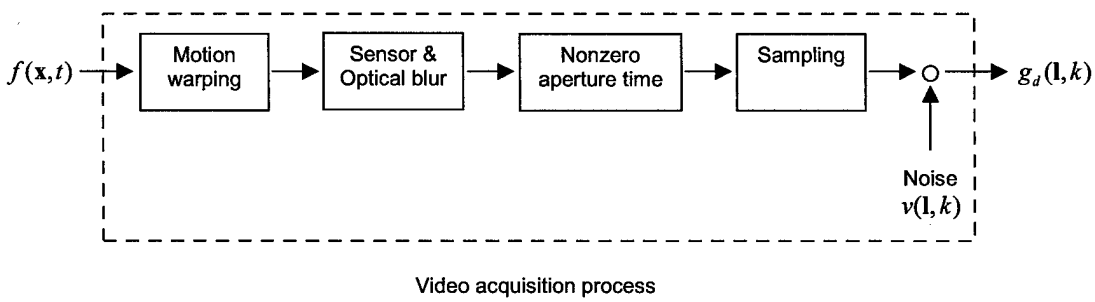


Figure 1: Video acquisition model.

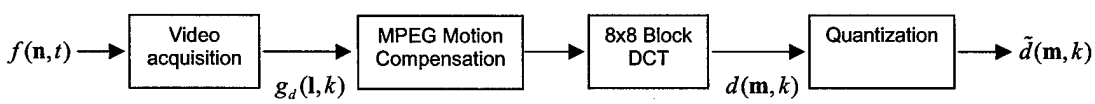


Figure 2: MPEG compression is appended to the video acquisition.

4.3 Fast Motion Estimation Using Low-bit-resolution Images

Reducing the temporal redundancy in image sequences, motion estimation is widely used in the video processing algorithms. Among different motion estimation techniques, block-matching algorithm (BMA) is the most popular one due to its high performance and low hardware cost. Block-matching algorithms find the displacement (motion vector) minimizing the matching difference between the reference block and the candidate blocks. When the search range is large, BMAs become computationally costly, which is a significant problem for real-time video processing applications. There are various fast motion estimation approaches proposed to reduce that computational cost. Unimodal error surface assumption techniques, multiresolution techniques, integral projection techniques, variable search range techniques, lower bit-resolution techniques are among them.

The focus of this report will be the low bit-resolution BMAs. In an early work, Gharavi and Mills [1] proposed a fast BMA based on one-bit quantization of the pixels. Natarajan *et al.* [2] made that algorithm even faster by using exclusive-OR (XOR) operation for the block-matching criterion.

Figure 4: A frame from the synthetically generated video sequence.

Although the XOR based block-matching criteria reduce the computational load significantly for one-bit resolution algorithms, the accuracy of the motion estimation is a big concern. Baek *et al.*

Figure 5: Bilinear interpolation.

[3] compared the performances of different bit-resolution images using reduced-bit mean absolute difference as the matching criterion. Lee *et al.* [4] presented a two-bit adaptive quantization scheme to increase the accuracy of the motion estimation. They also introduced a new block-matching criterion, Different Pixel Count (DPC), providing less hardware cost than the other reduced-bit matching criteria such as Pel Difference Classification [1] and Bit Truncation [3]. But all these block-matching criteria do not still come close to the efficiency of the simple XOR based methods.

In this work, we propose a new two-bit BMA that uses XOR operation as the correlation engine. Three-level quantization scheme is the key in this algorithm, enabling the use of XOR operation. The proposed algorithm is more accurate than the one-bit methods.

4.3.1 Three-level quantization of images

The block diagram of the proposed algorithm is given in Figure 11. The input images are first band-pass filtered and then quantized to three levels (Q3). The band-pass filter (BPF) is designed to remove the DC so that the overall method is less susceptible to the errors due to brightness

Figure 6: Proposed algorithm (8 frames used).

changes, and also remove the high-frequency noise. By quantizing the images to three levels, the accuracy of the motion estimation is increased in comparison to the two-level quantization methods while providing a fast implementation scheme based on XOR operation at the same time.

For the three-level quantization, the levels are represented by the two bit pairs: 10, 00 or 01, corresponding to the pixel values less than T1, between T1 and T2, and greater than T2, respectively, where T1 and T2 are the thresholds. Upon applying the XOR we have the following combinations and results:

$$00 \text{ XOR } 00 = 00,\ 10 \text{ XOR } 10 = 00,\ 01 \text{ XOR } 01 = 00$$
$$00 \text{ XOR } 01 = 01,\ 00 \text{ XOR } 10 = 10,\ 10 \text{ XOR } 01 = 11$$

From these sets of possible comparisons, we see that the number of 1 bits in the result can be a measure of correlation. It gives zero 1 bits if the pixels are highly correlated (00 XOR 00, 10 XOR 10, etc.), two 1 bits if the correlation is the lowest (10 XOR 01), and one 1 bit in between (00 XOR 01, 00 XOR 10). Therefore, the algorithm will count the number of 1 bits in the results of the XOR operation between two blocks, and choose the block with the least number of 1 bits as the matched

Figure 8: A frame from the synthetically generated video sequence.

block.

Figure 9: Bilinear interpolation.

4.3.2 Band-pass filter implementation

The band-pass filter operation is designed to achieve a fast and effective implementation. As depicted in Figure 12, the image is filtered by a 3×3 window of ones, and a 7×7 window of ones. The difference of these two filtered images gives the band-pass filtered image. With this kind of scheme, the filtering operation becomes extremely efficient: Without any multiplication, the pixels under the 3×3 and 7×7 windows are summed up, and then scaled once at the end. This scaling can also be simplified if the the window sizes are chosen to be factors of two, in which case the pixel values are scaled by bitwise shifting to the right.

4.3.3 Threshold selection

Since the overall method is aimed to be fast, the threshold selection technique is chosen to have low computational complexity. The simplest way of choosing a threshold is to choose a fixed threshold, that can be implemented by bit truncation. However, it is obvious that with this kind of threshold selection, the motion vectors cannot be computed accurately. Another simple method is to find the

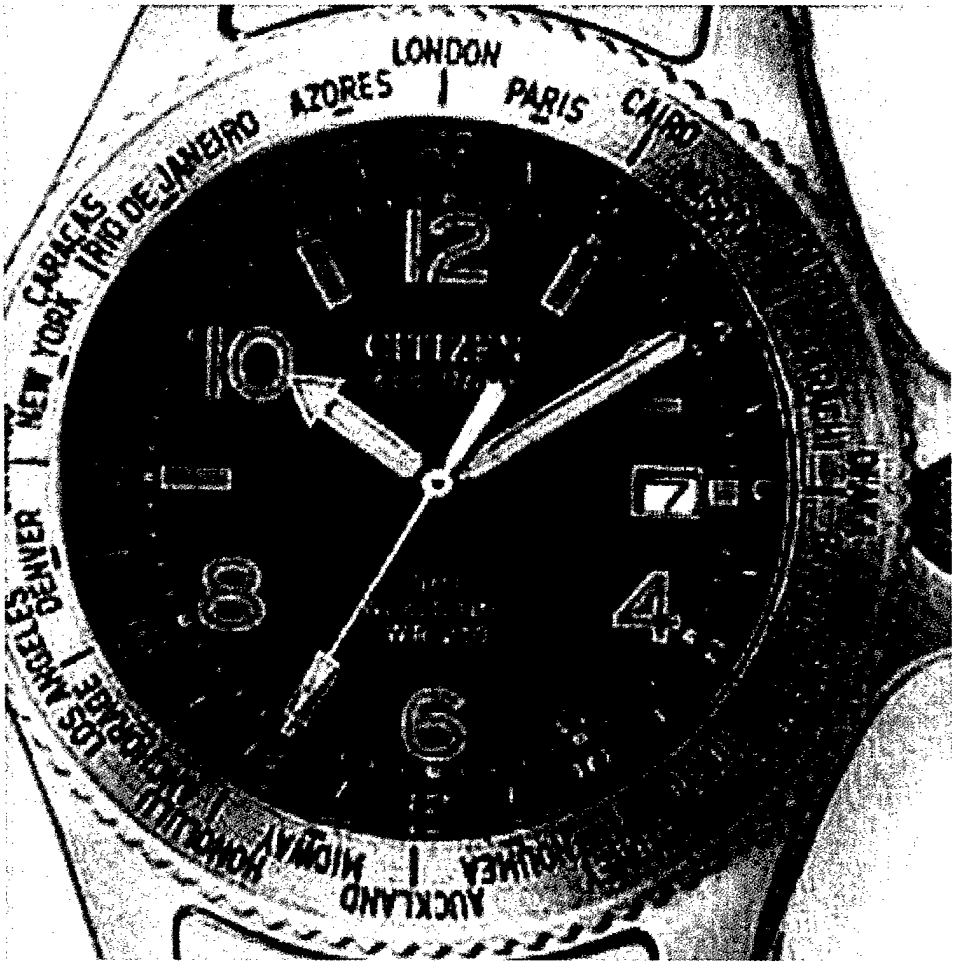


Figure 10: Proposed algorithm (8 frames used).

mean of the image, and then use a fixed offset value from the mean to determine T1 and T2. In other words, a fixed number is subtracted from the mean of the image to determine T1, and added to the mean to determine T2. We have performed experiments for different video sequences to determine an optimum threshold value. Figure 13 is a graph showing the average PSNR values for different quantization levels and band-pass filters, as a function of threshold values. In that figure, *Filter 3-5* indicates that windows of 3×3 and 5×5 are used in the band-pass filter implementation. Similarly, *Filter 3-7* indicates that windows of 3×3 and 7×7 are used. An important observation in the figure is that three-level quantization has better performance than two-level quantization at small threshold values. However, as the threshold value increases, the performance of the three-level quantization becomes worse than two-level quantization. Although this is the plot for a particular sequence ("Foreman" sequence), the same behaviour has been been observed for all video sequences tested. In our experiments, we have found that an offset value of one works sufficiently for the bandpass filter *Filter 3-7*, mentioned above.

4.3.4 Correlation surface search

The most important element in this scheme is the Packed Correlation Surface Search. As explained before, the correlation operation of an image region taken against another image region at some shift can be computed utilizing the XOR operation on packed words. For example, if a given processor has 32-bit data words, 16 pixels from one image line can be correlated against 16 pixels from another image line utilizing only a simple XOR operation. The correlation measure is simply the sum of the number of 1 bits in the resultant 32-bit word. By segmenting each line of the selected block into 16 pixel chunks, the correlation over the entire block, at a given shift, can be computed with a minimum number of operations. If two bits are used to represent each pixel, then two 32-bit data words can be used to compute the correlation for 16 pixels.

4.3.5 Generalization to higher-level quantization

It should be noted that in two-bit per pixel representation there are actually four possible quantization levels. However, the only way to use the sum of the number of 1 bits in the XORed registers as a distance metric is to use three different quantization levels. This can be generalized to higher levels as well. For instance, when the pixels are represented by three bits, there are four quantization levels possible if we want to use the sum of the number of 1 bits as the distance metric. These levels can be represented with 000, 001, 011, and 111. In this case, the distance between two intensities can be 0, 1, 2, or 3.

4.3.6 Experimental results

We have tested the proposed algorithm for various video sequences, including the standard "Susie", "Foreman", "Coast Guard", "Tennis", and "Flower Garden" sequences. The results are tabulated in Table 1. The size of each frame for the "Foreman" sequence is 288×352, for the others it is 240×352. We have used two-level and three-level quantization schemes, and compared it with the conventional 8-bit algorithm. The motion vectors are computed for every 8×8 or 16×16 blocks, and the numbers in parenthesis show this periodicity. As seen on the table, two- and three-level quantization algorithms have significantly lower search time than the standard 8-bit algorithm at the cost of 0.5 dB loss in PSNR. Three-level quantization is 0.2 dB better than the two-level quantization on the average. This 0.2 dB is a critical amount considering the fact that 8-bit algorithm is 0.5 dB better than reduced-bit algorithms on the average.

The search time results were obtained on a Pentium III 900MHz processor. We can also have an approximate picture of the computational complexities in terms of the number of operations per pixel. For the filtering operation, there are 59 flops (addition or multiplication) required for each pixel. (9 for the 3×3 window filter, 49 for the 7×7 filter, and 1 for the final subtraction.) For the three-level quantization operation, less than 1 flop (addition or multiplication) is required per pixel. And for the correlation search, only 65 flops per pixel is necessary for a search range of [-8, 7] pixels (full search). The total number of flops per pixel is only 125. This is a significant reduction when compared to the conventional 8-bit full search method, which requires more than 2700 flops per pixel.

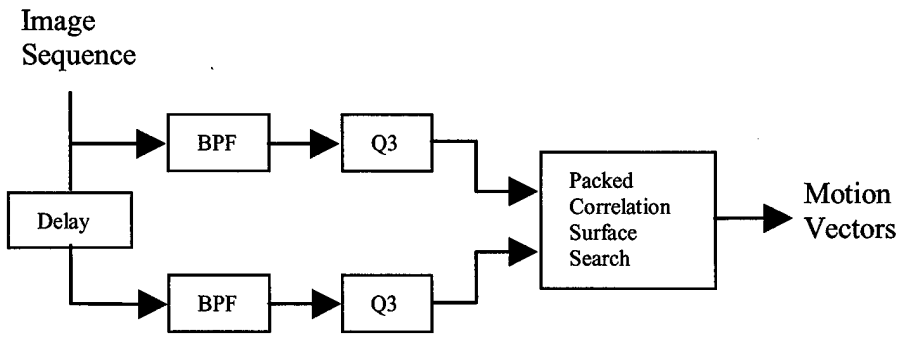


Figure 11: Block diagram of the proposed BMA

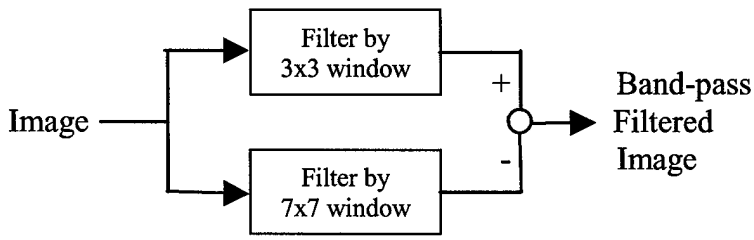


Figure 12: Computation of the band-pass filtered image

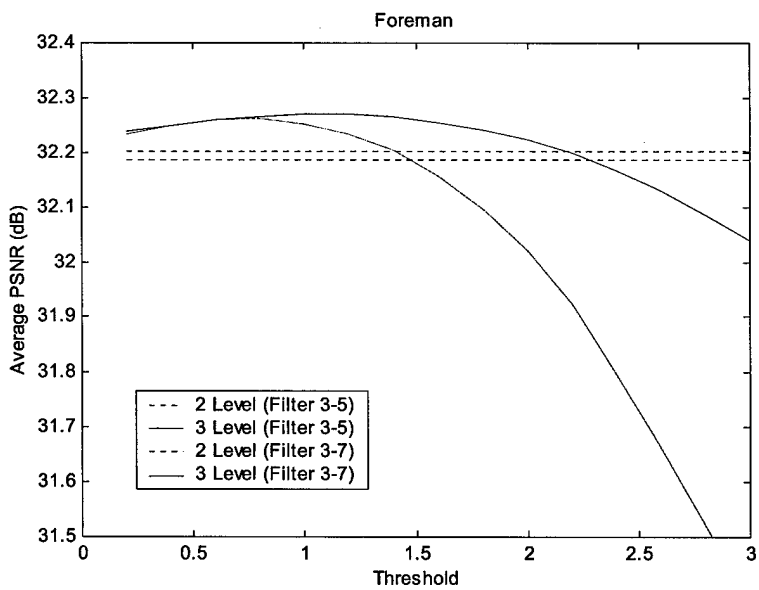


Figure 13: Threshold selection

4.4 Multi-Frame Gray-Scale Resolution Enhancement

When images are digitized, a certain number of bits is assigned to each pixel to represent its intensity. The number of bits, the *bit depth*, determines the number of gray levels between the minimum and maximum values that the imaging device can capture. There will be a loss of gray-scale resolution if the bit depth is not sufficient. When a set of low bit-depth images is available

Table 1: Average PSNR and Search Time Comparison.

Method	Video Sequence									
	Susie (240x352)		Foreman (288x352)		Coast Guard (240x352)		Tennis (240x352)		Flower (240x352)	
	PSNR (dB)	Time (sec)	PSNR (dB)	Time (sec)	PSNR (dB)	Time (sec)	PSNR (dB)	Time (sec)	PSNR (dB)	Time (sec)
2 Level (8)	36.0247	0.469	32.8750	0.562	27.5283	0.469	30.6896	0.469	25.1050	0.469
3 Level (8)	36.0165	0.719	32.9885	0.875	27.6164	0.719	31.0983	0.719	25.2322	0.719
2 Level (16)	35.8953	0.250	31.5188	0.312	27.4410	0.250	29.6036	0.250	24.3092	0.250
3 Level (16)	35.7577	0.313	31.7258	0.375	27.5311	0.313	30.1381	0.313	24.6414	0.313
8-bit (16)	36.1092	47.1	32.8216	47.1	27.9828	47.1	31.4615	47.1	25.9193	47.1

that are slightly different from each other because of motion or illumination, their non-redundant information can be combined to enhance the gray-scale resolution. We refer to this process of multi-frame gray-scale resolution enhancement as *superprecision.*

Superprecision can be applied in several application areas. One of the most important of these is in medical imaging. With these images low-contrast details are often extremely critical for diagnosis, but when the bit depth is insufficient, these details may be lost. Superprecision reconstruction has the potential to regain these details by combining the non-redundant information that is present in a set of images. Military automatic target detection, aerial and satellite remote sensing, and high-quality scanning applications can also use make effective use of superprecision reconstruction to enhance gray-scale resolution. The conversion of images from a low bit-depth format to a higher one (e.g., conversion from 8-bit GIF to 24-bit JPEG) is also a potential application of this technology.

The superprecision problem is very similar to the superresolution problem in which higher spatial resolution is sought from a set of low spatial resolution images [63]. Although the superresolution problem has received a considerable amount of attention, the superprecision has not been as actively researched. In an early paper, Cheeseman *et al.* [16] proposed increasing both the spatial and gray-scale resolution at the same time by using a maximum *a posteriori* probability estimator. They assumed a Gaussian model for all of the probability distributions and used Jacobi's method to solve the problem iteratively. In this paper, we propose a deterministic method based on a projections onto convex sets (POCS) technique that does not make any assumptions about the underlying probability densities. Using this method, an initial high-resolution image estimate is projected onto constraint sets that are derived from the low resolution gray-scale image observations. The method can work either in the spatial domain or in the transform domain, where it is possible to include details of the compression process. In some cases superprecision and superresolution can be achieved at the same time.

Section 2 presents the imaging model to be used in the reconstruction. The POCS-based superprecision approach is presented in Section 3, and Section 4 gives the final algorithm. Section 5 presents some experimental results, and Section 6 concludes the paper.

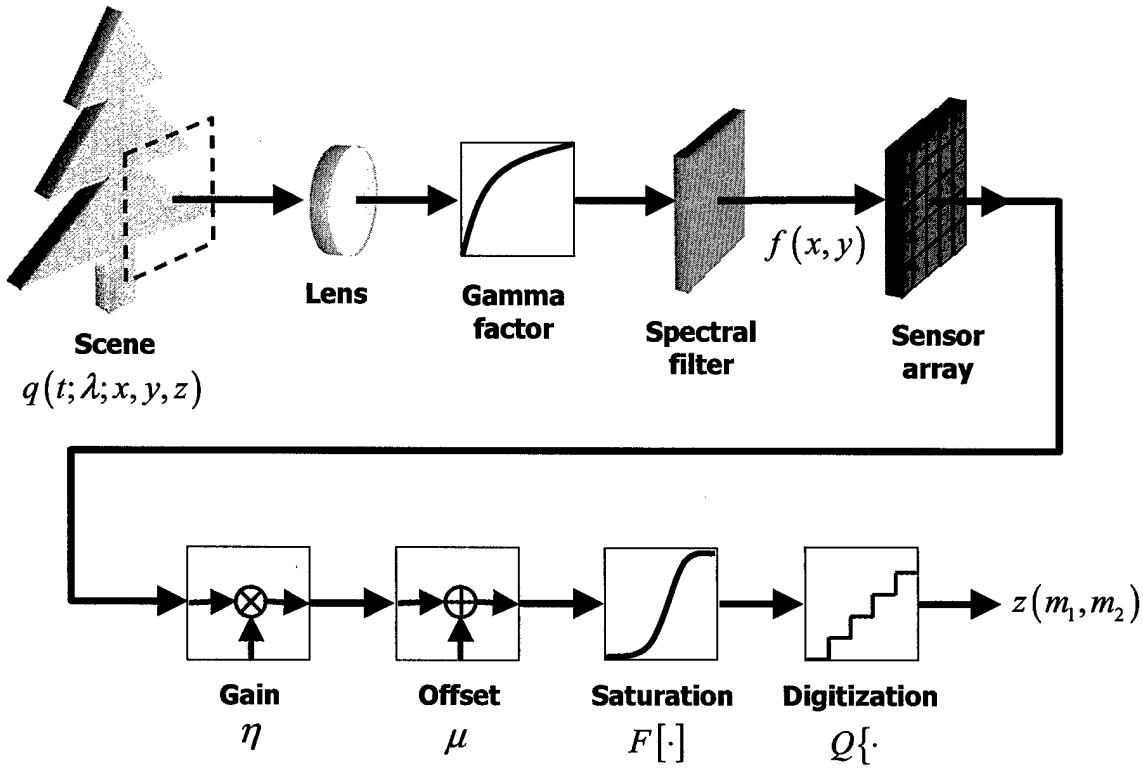


Figure 14: A model for the production of a low-resolution image from a high-precision, high-resolution image by the image recording process.

4.4.1 The Imaging Model

This section presents a model that establishes the connection between a high-resolution source image and multiple low bit-depth image observations. This connection will be used in the next section to enhance the gray-scale resolution of the images.

As seen in Figure 39, the model has two components. The first models the image capture process. A high-resolution image, $f^{(N_1)}(n_1, n_2)$, is captured by an imaging device to produce the low resolution images, $g^{(N_1)}(m_1, m_2, k)$. Here the superscript N_1 represents the number of bits used to represent each pixel and (n_1, n_2) and (m_1, m_2, k) are the spatial pixel coordinates of the high-resolution image and the kth low-resolution image, respectively. The image capture process is a linear, shift-varying (LSV) operation that includes motion (of the camera or the objects in the scene), blur (because of the nonzero sensor aperture time, the nonzero physical dimensions of the individual sensor elements, the degree of focus, etc.), and sampling with a low-resolution grid [42]. In this paper, we model all of these effects except for the sensor aperture time, which is taken to be zero. According to this model, the mapping from a high-resolution image to a low spatial-resolution image is expressed as a weighted sum of the high-resolution image pixels, where the weights are the values of a space-invariant point-spread function (PSF) at the corresponding pixel locations. The center of the PSF depends upon the motion between the high-resolution image and the low-resolution images. This is depicted in Figure 15. Motion vectors from each low spatial-resolution image to the high-resolution image determine how each pixel is mapped. The normalized PSF that characterizes the camera is centered at that location, and from it the weights on the high-resolution image grid are found.

Defining $h(m_1, m_2, k; n_1, n_2)$ as this mapping, we can model the image capture process by:

$$g^{(N_1)}(m_1, m_2, k) = \sum_{n_1, n_2} h(m_1, m_2, k; n_1, n_2) f^{(N_1)}(n_1, n_2). \tag{8}$$

Equation 8 provides the relation between a single high-resolution image and the set of low spatial-resolution images. Further details concerning this modeling can be found in [46].

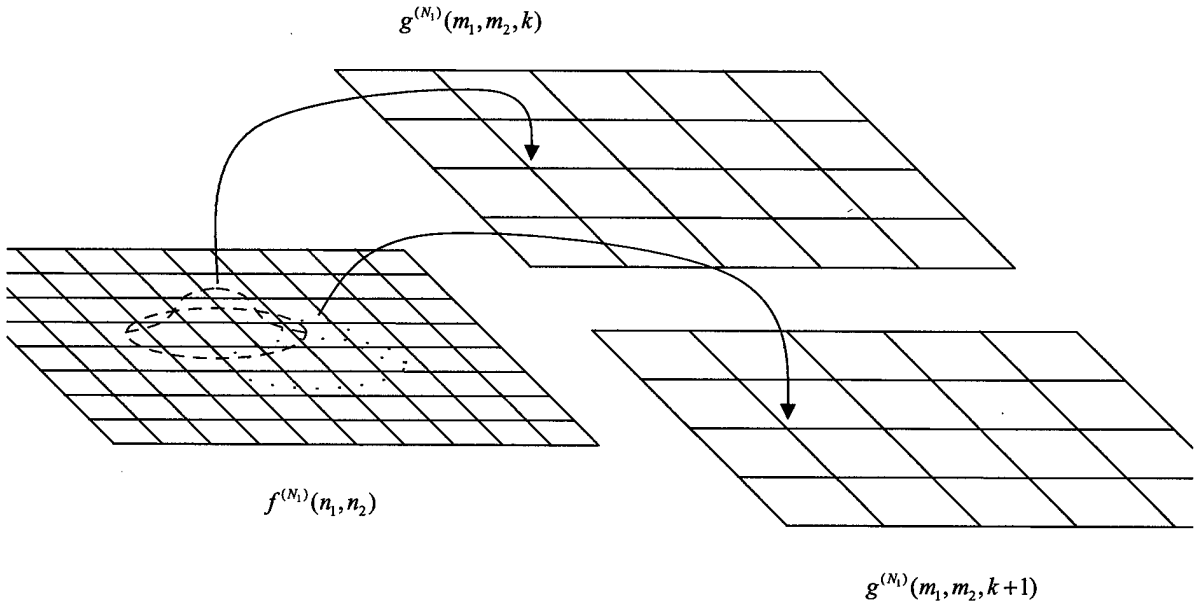


Figure 15: A mapping of the high resolution point-spread function to the lower resolution sampling lattices.

The second sub-block models the gray-scale resolution reduction, which is nothing but a reduction in the number of bits used to represent each pixel. Bit depth reduction from N_1 bits to N_2 bits is equivalent to the operation:

$$g^{(N_2)}(m_1, m_2, k) = \text{round}\left\{2^{N_2 - N_1} g^{(N_1)}(m_1, m_2, k)\right\}, \tag{9}$$

where $g^{(N_2)}(m_1, m_2, k)$ corresponds to the low-resolution (both in gray-scale and in the spatial variables) image observations. The factor $2^{N_2 - N_1}$ scales the image to the new range, and the rounding operation $round\{\cdot\}$, which rounds the argument to the nearest integer, effects the quantization.

Letting $\delta(m_1, m_2, k)$ denote the error introduced by rounding, we can rewrite Equation 9 as

$$g^{(N_2)}(m_1, m_2, k) = 2^{N_2 - N_1} g^{(N_1)}(m_1, m_2, k) + \delta(m_1, m_2, k), \tag{10}$$

where

$$|\delta(m_1, m_2, k)| < 0.5. \tag{11}$$

If the images are transform coded (using JPEG or MPEG for instance), a similar relation can be formulated in the transform domain. The most common transform is the discrete cosine transform (DCT) applied on 8×8 blocks. Taking an 8×8 block-DCT of Equation 10, we get:

$$G^{(N_2)}(l_1, l_2, k) = 2^{N_2 - N_1} G^{(N_1)}(l_1, l_2, k) + \Delta(l_1, l_2, k), \tag{12}$$

where $G^{(N_2)}(l_1, l_2, k)$, $G^{(N_1)}(l_1, l_2, k)$, and $\Delta(l_1, l_2, k)$ are the DCTs of $g^{(N_2)}(m_1, m_2, k)$, $g^{(N_1)}(m_1, m_2, k)$, and $\delta(l_1, l_2, k)$, respectively. $\Delta(l_1, l_2, k)$ is also bounded in magnitude since $\delta(l_1, l_2, k)$ is bounded, but its bounds depend on the position (l_1, l_2). They are derived in the Appendix. (It should be noted that the bounds of $\Delta(l_1, l_2, k)$ are *not* equal to the DCT of the bounds of $\delta(l_1, l_2, k)$.) Applying the block-DCT to Equation 8, we can write $G^{(N_1)}(l_1, l_2, k)$ as:

$$G^{(N_1)}(l_1, l_2, k) = \sum_{n_1, n_2} h_{DCT}(l_1, l_2, k; n_1, n_2) f^{(N_1)}(n_1, n_2), \tag{13}$$

where $h_{DCT}(l_1, l_2, k; n_1, n_2)$ is the block-DCT of $h(m_1, m_2, k; n_1, n_2)$. Mathematically, the relation can be written as:

$$h_{DCT}(l_1, l_2, k; n_1, n_2) = \sum_{m_1=L(l_1)}^{L(l_1)+7} \sum_{m_2=L(l_2)}^{L(l_2)+7} K\left(l_1, l_2; (m_1)_8, (m_2)_8\right) h(m_1, m_2, k; n_1, n_2), \tag{14}$$

where $(\cdot)_8$ is the *modulo* 8 operator, $L(x) \equiv \lfloor x/8 \rfloor$, and $K(l_1, l_2; m_1, m_2)$ is the DCT kernel given by:

$$K(l_1, l_2; m_1, m_2) = k_{l_1} k_{l_2} \cos\left(\frac{(2m_1+1)l_1\pi}{16}\right) \cos\left(\frac{(2m_2+1)l_2\pi}{16}\right). \tag{15}$$

k_{l_i} is the normalization factor

$$k_{l_i} = \begin{cases} \frac{1}{4\sqrt{2}} & , \quad l_i \neq 0 \\ \frac{1}{4} & , \quad l_i = 0 \end{cases} \quad , \qquad i = 1, 2. \tag{16}$$

To summarize we have two equations that relate the HR image to LR images and DCT coefficients:

$$g^{(N_2)}(m_1, m_2, k) = 2^{N_2 - N_1} \sum_{n_1, n_2} h(m_1, m_2, k; n_1, n_2) f^{(N_1)}(n_1, n_2) + \delta(m_1, m_2, k), \tag{17}$$

and

$$G^{(N_2)}(l_1, l_2, k) = 2^{N_2 - N_1} \sum_{n_1, n_2} h_{DCT}(l_1, l_2, k; n_1, n_2) f^{(N_1)}(n_1, n_2) + \Delta(l_1, l_2, k). \tag{18}$$

These two equations will be used in the POCS-based superprecision estimation algorithms. Equation 18 will also be extended to include the compression process where the DCT coefficients are quantized according to quantization tables.

4.4.2 Superprecision Methods

This section formulates POCS-based superprecision in the spatial and transform domains. The transform-domain formulation also includes the quantization process that is common in image and video compression standards such as JPEG, MPEG, and H.263.

Spatial-Domain POCS Solution

From Equation 17, it is seen that the value of $2^{N_2 - N_1} g^{N_1)}(m_1, m_2, k)$ falls within the 0.5 proximity of the observed pixel intensity $g^{(N_2)}(m_1, m_2, k)$. This information can be used to define a convex constraint set for each observed pixel. The method follows.

The image capture process shown in Figure 39 is applied on an initial high-resolution image estimate $x^{(N_1)}(n_1, n_2)$, and then scaled by the factor $2^{N_2-N_1}$. The result is compared to the observations $g^{(N_2)}(m_1, m_2, k)$. It is known that the residual, the difference between the computed and the observed images, must be less than 0.5 in magnitude if the estimate is correct. If the residual is greater than 0.5, then the error is back-projected onto the initial estimate so that the next time the image capture model and scaling are applied, it will fall within the 0.5 proximity of the observations.

Defining the residual as:

$$r_x(m_1, m_2, k) \equiv g^{(N_2)}(m_1, m_2, k) - 2^{N_2-N_1} \sum_{n_1,n_2} h(m_1, m_2, k; n_1, n_2) x^{(N_1)}(n_1, n_2), \tag{19}$$

we can write the convex constraint set for an arbitrary image $x^{(N_1)}(n_1, n_2)$ as follows:

$$C_{(m_1,m_2,k)} = \left\{ x^{(N_1)}(n_1, n_2) : |r_x(m_1, m_2, k)| < 0.5 \right\}. \tag{20}$$

The projection operation onto these convex sets is given by:

$$P_{(m_1,m_2,k)} \left[x^{(N_1)}(n_1, n_2) \right] = \begin{cases} x^{(N_1)}(n_1, n_2) + \frac{2^{N_1-N_2}(r_x(m_1,m_2,k)-0.5)h(m_1,m_2,k;n_1,n_2)}{\sum_{n_1,n_2} |h(m_1,m_2,k;n_1,n_2)|^2}, & r_x(m_1, m_2, k) > 0.5 \\ x^{(N_1)}(n_1, n_2), & -0.5 \le r_x(m_1, m_2, k) \le 0.5 \\ x^{(N_1)}(n_1, n_2) + \frac{2^{N_1-N_2}(r_x(m_1,m_2,k)+0.5)h(m_1,m_2,k;n_1,n_2)}{\sum_{n_1,n_2} |h(m_1,m_2,k;n_1,n_2)|^2}, & r_x(m_1, m_2, k) < -0.5 \end{cases} \tag{21}$$

Compressed-Domain POCS Solution

Equation 18 provides the relation between the DCT coefficients $G^{(N_2)}(l_1, l_2, k)$ and the HR image $f^{(N_1)}(n_1, n_2)$. The DCT coefficients $G^{(N_2)}(l_1, l_2, k)$ are quantized in the process of compression. Defining $\tilde{G}^{(N_2)}(l_1, l_2, k)$ as the quantized DCT coefficients, and $Q(l_1, l_2, k)$ as the quantization error, we can write:

$$\tilde{G}^{(N_2)}(l_1, l_2, k) = G^{(N_2)}(l_1, l_2, k) + Q(l_1, l_2, k), \tag{22}$$

where $Q(l_1, l_2, k)$ is bounded by half of the quantization step size at location (l_1, l_2). Using Equation 18, we get

$$G^{(N_2)}(l_1, l_2, k) = 2^{N_2-N_1} \sum_{n_1,n_2} h_{DCT}(l_1, l_2, k; n_1, n_2) f^{(N_1)}(n_1, n_2) + \Delta(l_1, l_2, k) + Q(l_1, l_2, k). \tag{23}$$

The sum of of $\Delta(l_1, l_2, k)$ and $Q(l_1, l_2, k)$ is bounded by $B(l_1, l_2, k)$, which is equal to the sum of the bounds of $\Delta(l_1, l_2, k)$ and $Q(l_1, l_2, k)$. That is,

$$|\Delta(l_1, l_2, k) + Q(l_1, l_2, k)| < B(l_1, l_2, k), \tag{24}$$

where

$$B(l_1,l_2,k) = 0.5 \sum_{m_1=L(l_1)}^{L(l_1)+7} \sum_{m_2=L(l_2)}^{L(l_2)+7} \left|K\left(l_1,l_2;(m_1)_8,(m_2)_8\right)\right| + \frac{\textit{Quantization step size at location } (l_1,l_2)}{2}. \quad (25)$$

The first term on the right-hand side of Equation 25 is derived in the Appendix. Equation 23 along with the bound $B(l_1,l_2,k)$ allow us to write a POCS reconstruction algorithm in the compressed domain that is analogous to the spatial-domain algorithm derived in the previous section.

Now we can follow the same procedure as in the previous section to create the projection operator. Defining the compressed-domain residual as

$$R_x(l_1,l_2,k) \equiv \tilde{G}^{(N_2)}(l_1,l_2,k) - 2^{N_2-N_1} \sum_{n_1,n_2} h_{DCT}(l_1,l_2,k;n_1,n_2) x^{(N_1)}(n_1,n_2), \quad (26)$$

we can write the convex constraint set for an arbitrary image $x^{(N_1)}(n_1,n_2)$ as follows:

$$C_{(l_1,l_2,k)} = \left\{ x^{(N_1)}(n_1,n_2) : |R_x(l_1,l_2,k)| < B(l_1,l_2,k) \right\}. \quad (27)$$

The projection operation onto these convex sets is given by:

$$P_{(l_1,l_2,k)}\left[x^{(N_1)}(n_1,n_2)\right] =$$

$$\begin{cases} x^{(N_1)}(n_1,n_2) + \frac{2^{N_1-N_2}(R_x(l_1,l_2,k)-B(l_1,l_2,k))h_{DCT}(l_1,l_2,k;n_1,n_2)}{\sum_{n_1,n_2}|h_{DCT}(l_1,l_2,k;n_1,n_2)|^2}, & R_x(l_1,l_2,k) > B(l_1,l_2,k) \\ x^{(N_1)}(n_1,n_2), & -B(l_1,l_2,k) \le R_x(l_1,l_2,k) \le B(l_1,l_2,k) \\ x^{(N_1)}(n_1,n_2) + \frac{2^{N_1-N_2}(R_x(l_1,l_2,k)+B(l_1,l_2,k))h_{DCT}(l_1,l_2,k;n_1,n_2)}{\sum_{n_1,n_2}|h_{DCT}(l_1,l_2,k;n_1,n_2)|^2}, & R_x(l_1,l_2,k) < -B(l_1,l_2,k) \end{cases} \quad (28)$$

4.4.3 Implementation

Implementations of both the spatial- and transform-domain algorithms are very similar. In the spatial-domain implementation the error back-projected is the error in the pixel intensities. In the transform-domain implementation it is the error in the DCT coefficients that is back-projected. An illustration of using multiple constraint sets to restrict the solution set is illustrated in Figure 16. The final algorithm for spatial-domain (and compressed-domain) reconstruction is the follows.

1. Choose a reference frame and bilinearly interpolate it to get an initial fine-grid image.
2. Extend the bit depth of this image to the required bit depth by multiplying each pixel intensity by $2^{N_1-N_2}$ and filling in the last $N_1 - N_2$ bits randomly. The resulting image, $x^{(N_1)}(n_1,n_2)$, is an initial estimate for the high resolution image $f^{(N_1)}(n_1,n_2)$.
3. Compute the motion between the reference frame and one of the low-bit-depth images, $g^{(N_2)}(m_1,m_2,k)$.

4. Using the motion estimates, compute the mapping $h(m_1, m_2, k; n_1, n_2)$ for each pixel in the current image $g^{(N_2)}(m_1, m_2, k)$. (Compute $h_{DCT}(l_1, l_2, k; n_1, n_2)$ for each DCT coefficient in the case of transform-domain implementation)

5. For each pixel (DCT coefficient) in the current image,

   (a) Compute the pixel intensity (DCT coefficient value) from the estimate $x^{(N_1)}(n_1, n_2)$ by applying the image capture processing model and scaling,

   (b) Compute the residual $r_x(m_1, m_2, k)$ ($R_x(l_1, l_2, k)$) and back-project the error to the estimate $x^{(N_1)}(n_1, n_2)$ using Equation 21 (Equation 26).

6. Stop, if a stopping criterion is reached; otherwise, choose another low-bit-depth image, and go to step 2.

It should be noted that, by construction, this algorithm has the potential to achieve both spatial and gray-scale resolution enhancement at the same time. If the high-resolution image estimate has a finer grid than the observations, both spatial and gray-scale resolution enhancement are achieved. If they have the same sampling grid, only gray-scale resolution enhancement is achieved.

4.4.4 Experimental Results

The high-resolution image MONEY, shown in Figure 17 is jittered, blurred by a Gaussian kernel (with a support of 5×5 and variance of 2.5), and downsampled to produce eight spatially low-resolution images. Their gray-scale resolution is then reduced from eight bits to a lower number of bits. The lower-precision images are then processed by the proposed algorithm to produce high-quality images with twice the spatial resolution and eight bit gray-scale resolution. Figures 18 and 20 show four- and three-bit MONEY images. The reconstructed images are shown in Figures 19 and 21. The MONEY images are also MPEG-1 encoded (using eight frames) to illustrate the transform-domain implementation. Figures 22 and 23 depict the reconstructed images.

The reconstructions algorithm is also applied to images that have a bit depth of only one and two bits. Figures 24 through 27 show these cases.

The reconstructions made from four- and three-bit images show significant improvement in both gray-scale and spatial resolution. Although the image quality for the reconstructed images from a bit-depth of two and one bits are not satisfactory, characters that are not readable on the originals become readable after the reconstruction.

In the experiments, motion is computed using the Hierarchical Block Matching (HBM) method of Bierling [141]. Three hierarchical levels are used with the Mean Absolute Difference (MAD) as the matching criterion. In the final level, motion estimates are obtained with one-quarter pixel accuracy.

4.4.5 Conclusion

In this paper, we presented a superprecision method that increases the gray-scale resolution by combining the non-redundant information from a set of low bit-depth, low resolution images. It is based on the projections onto convex sets (POCS) technique where the convex sets are defined using the bit-depth reduction information. With the transform-domain implementation, the quantization error that results from compression can also be included in the reconstruction. The method can also increase the spatial resolution if a finer sampling grid is used for the initial high-resolution image estimate. Therefore, the proposed method can be considered as a generalization to superresolution reconstruction.

Appendix:

The error $\delta(m_1, m_2, k)$ is bounded by 0.5. When the image is transformed to the DCT domain, there will be a new error $\Delta(l_1, l_2, k)$. This appendix determines the bounds that are appropriate for $\Delta(l_1, l_2, k)$.

Following from Equations 10 and 12, the bound $\max|\Delta(l_1, l_2, k)|$ can be found by:

$$\begin{aligned} \max|\Delta(l_1, l_2, k)| &= \max\left| \begin{array}{c} DCT\left\{2^{N_2-N_1} g^{(N_1)}(m_1, m_2, k) + \delta(m_1, m_2, k)\right\} \\ - DCT\left\{2^{N_2-N_1} g^{(N_1)}(m_1, m_2, k)\right\} \end{array} \right| \\ &= \max|DCT\left\{\delta(m_1, m_2, k)\right\}| \end{aligned} \tag{29}$$

Using the definition of $DCT\{\cdot\}$, we get:

$$\begin{aligned} \max|\Delta(l_1, l_2, k)| &= \max\left| \sum_{m_1=L(l_1)}^{L(l_1)+7} \sum_{m_2=L(l_2)}^{L(l_2)+7} K\left(l_1, l_2; (m_1)_8, (m_2)_8\right) \delta(m_1, m_2, k) \right| \\ &= 0.5 \sum_{m_1=L(l_1)}^{L(l_1)+7} \sum_{m_2=L(l_2)}^{L(l_2)+7} |K\left(l_1, l_2; (m_1)_8, (m_2)_8\right)|. \end{aligned} \tag{30}$$

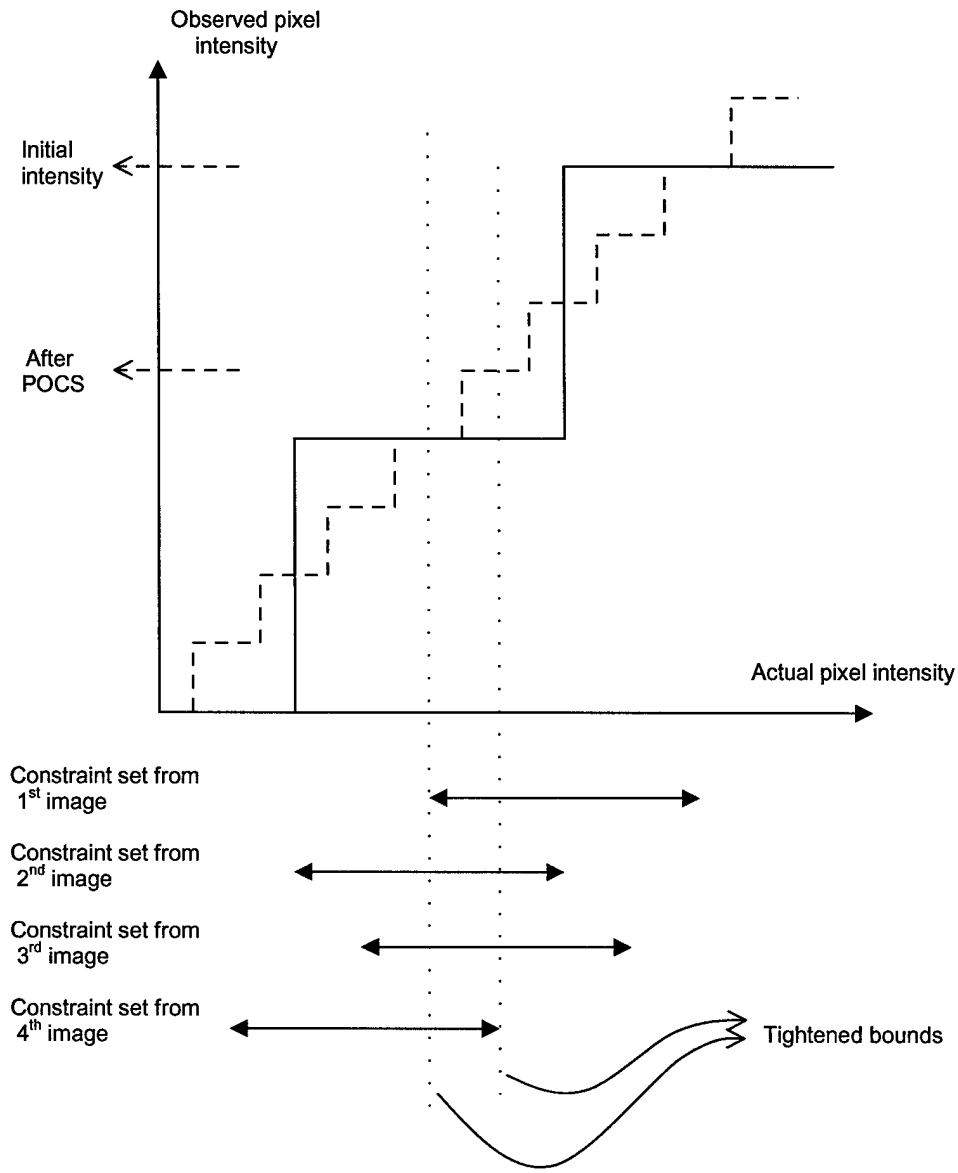


Figure 16: Illustration of how the POCS algorithm uses multiple coarse quantization values to produce a finer estimate of a sample value.

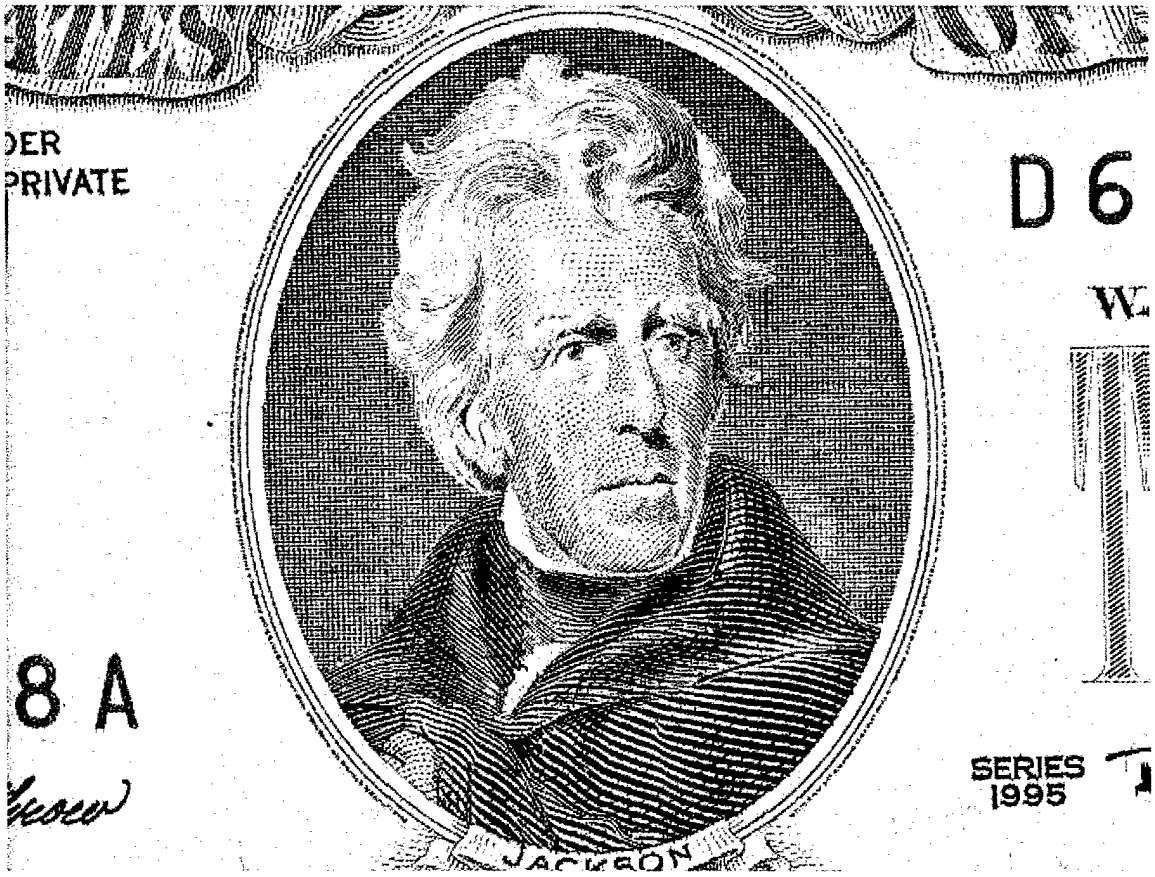


Figure 17: Original MONEY image.

4.5 Face Recognition from Video

The performance of existing face recognition systems decreases significantly if the resolution of the face image falls below a certain level. This is especially critical in surveillance imagery where often only a low-resolution video sequence of the face is available. If these low-resolution images are passed to a face recognition system, the performance is usually unacceptable. Therefore, super-resolution techniques have been proposed for face recognition that attempt to obtain a high-resolution face image by combining the information from multiple low-resolution images [15, 6, 39, 14]. In general, super-resolution algorithms try to regularize the ill-posedness of the problem using prior knowledge about the solution, such as smoothness or positivity [21]. Recently, researchers have proposed algorithms that attempt to use model-based constraints in regularization. While [15] demonstrates how super-resolution (without model-based priors) can improve the face recognition rate, [6], [39], and [14] provide super-resolution algorithms that use face-specific constraints for regularization.

All these systems propose super-resolution as a separate preprocessing block in front of a face recognition system. In other words, their main goal is to construct a high-resolution, visually improved face image that can later be passed to a face recognition system for improved performance. This is perfectly valid as long as computational complexity is not an issue. However, in a real-time surveillance scenario where the super-resolution algorithm is expected to work on continuous video

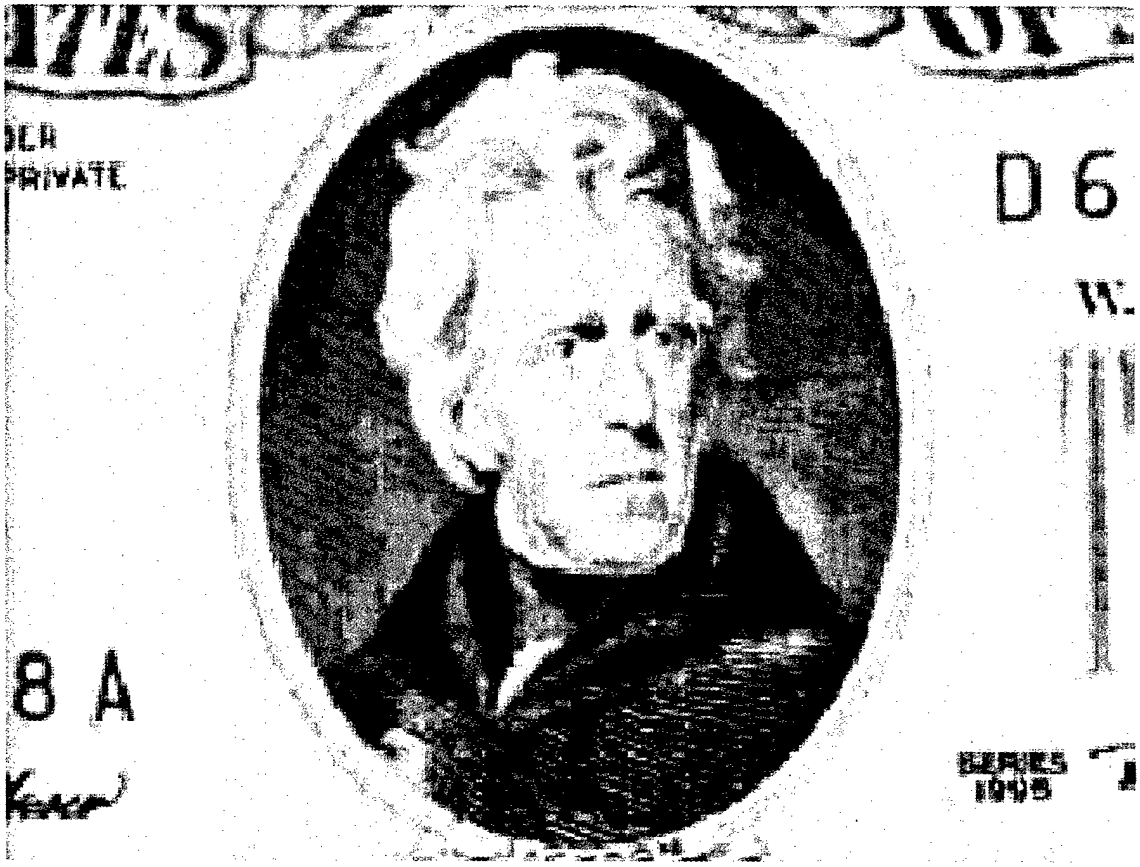


Figure 18: Low resolution four-bit MONEY image (one of eight).

streams, computational complexity is usually a very critical issue. In this paper, we propose an efficient super-resolution method for face recognition that embeds the super-resolution reconstruction into the face recognition system. This is based on the observation that nearly all state-of-the-art face recognition systems use some kind of front end dimensionality reduction, and that a lot of detailed information generated by a preprocessing type super-resolution algorithm is not used by the face recognition block. Hence, we propose to embed the super-resolution reconstruction into the low-dimensional framework of the face recognition system so that only the necessary information is reconstructed without any unnecessary overhead. In addition to the computational complexity reduction, we also derive face-specific constraints for the low-dimensional framework and demonstrate how they improve the performance.

Currently, by far the most popular dimensionality reduction technique in face recognition is to use subspace projections based on the Karhunen-Loeve Transform (KLT). This type of dimensionality reduction has been central to the development of face recognition algorithms for the last ten years. We propose to use a similar KLT-based dimensionality reduction technique to decrease the computational cost of the super-resolution algorithm by transforming it from a problem in the pixel domain to a problem in the lower-dimensional face subspace.

There are two important sources of noise in this problem. One is the observation noise that results from the imaging system. The other is the subspace representation error, which is a result of the

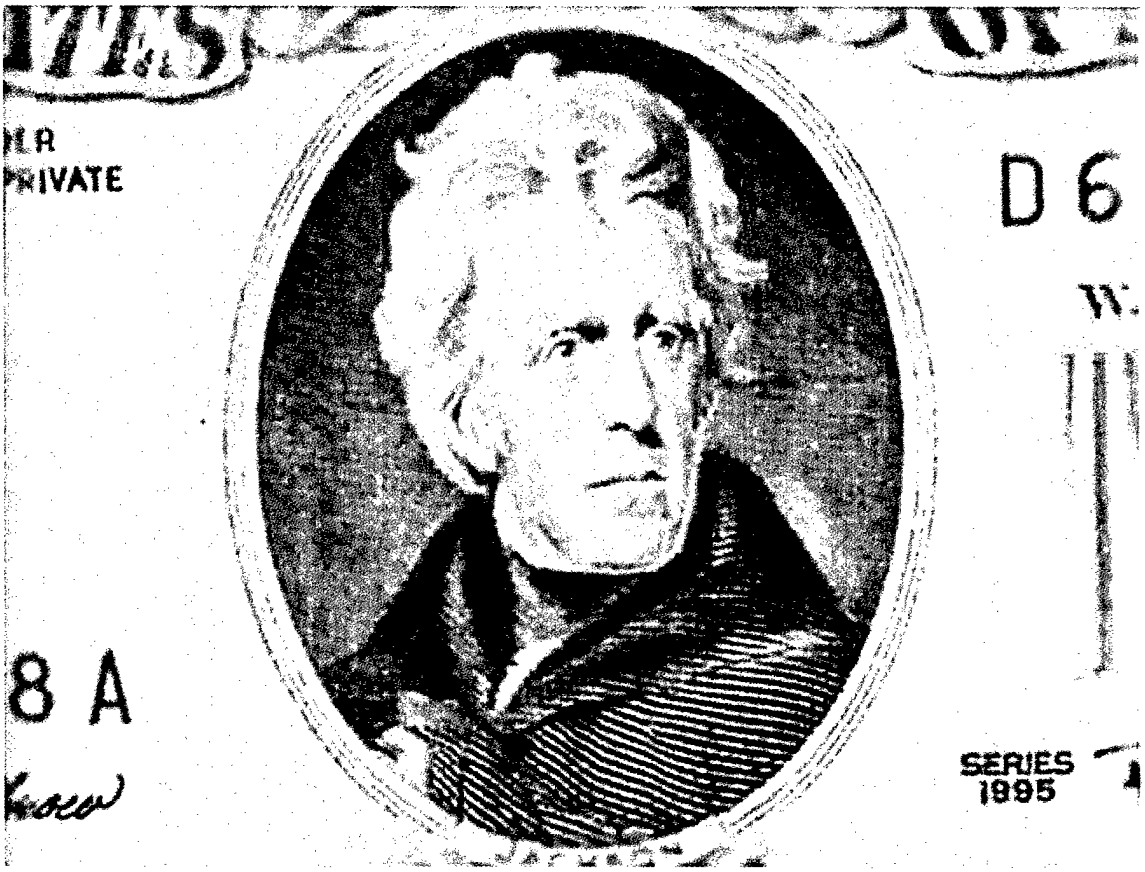


Figure 19: Reconstruction from the four-bit images shown in Figure 18.

dimensionality reduction. We derive the statistics of these noise processes *for the low-dimensional subspace* by using the examples from the human face image class. Substitution of this model-based information into the algorithm provides a higher robustness to noise. We test our system on both real and synthetic video sequences.

In Section 2, we briefly review the KLT-based dimensionality reduction method for face recognition. Then, in Section 3, we formulate the super-resolution problem in the low-dimensional framework. Section 4 details the reconstruction algorithm, and Section 5 provides experimental results. Conclusions are given in Section 6.

4.5.1 Dimensionality reduction for face recognition

Mathematically, the eigenface method tries to represent a face image as a linear combination of orthonormal vectors, called eigenfaces. These eigenfaces are obtained by finding the eigenvectors of the covariance matrix of the training face image set [68]. The eigenvectors corresponding to the largest L eigenvalues span a linear subspace that can reconstruct the face images with minimum reconstruction error in the least squares sense. This L-dimensional subspace is called the face space. Assuming $\mathbf{x}$ is the lexicographically ordered face image and $\mathbf{\Phi}$ is the matrix that contains

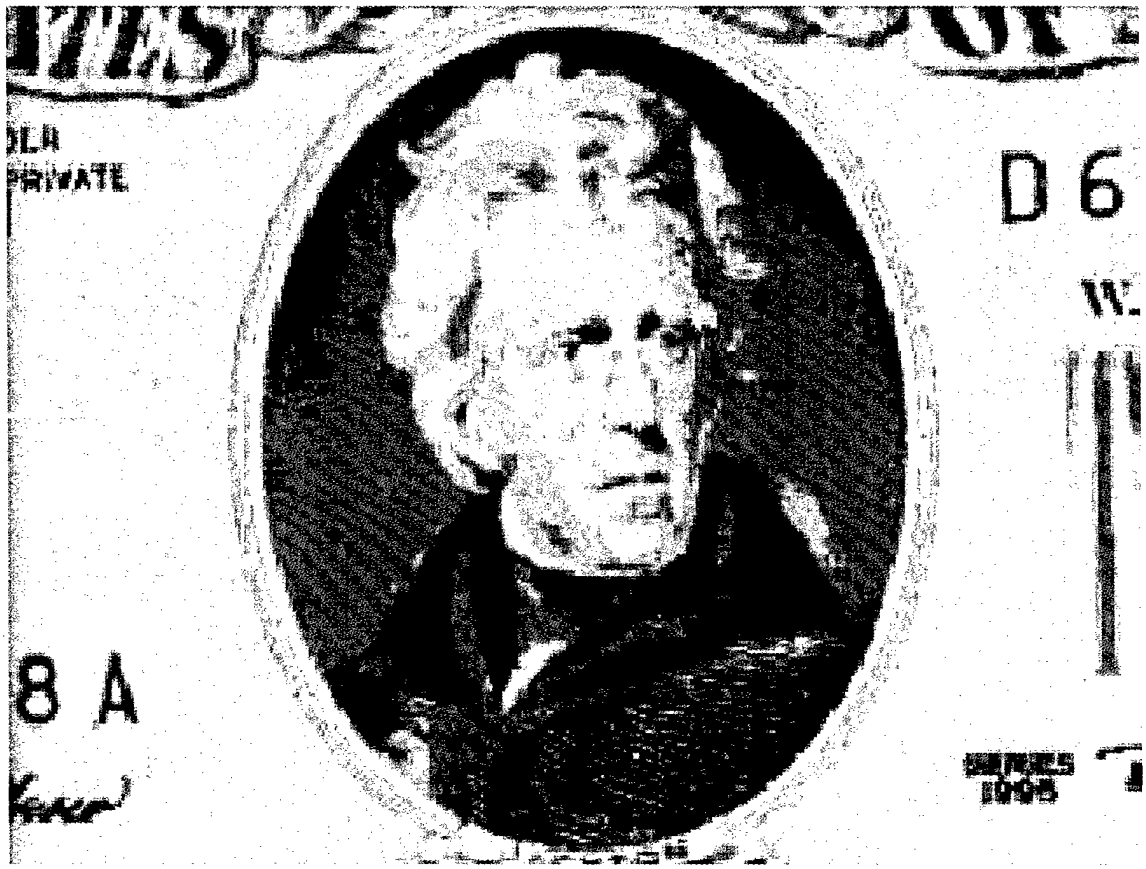


Figure 20: Low resolution three-bit MONEY image (one of eight).

the eigenfaces as its columns, we can write:

$$\mathbf{x} = \mathbf{\Phi a} + \mathbf{e_x} \tag{31}$$

where $\mathbf{a}$ is the $L \times 1$ feature vector that represents the face, and $\mathbf{e_x}$ is the subspace representation error for the face image.

4.5.2 Super-resolution in the face subspace

In this section, we formulate the super-resolution problem in the low-dimensional face subspace. In such a formulation, the observations are inaccurate feature vectors of a subject, and the reconstruction algorithm estimates the true feature vector. We start with the observation model for pixel-domain super-resolution, and then derive the observation model for face-space super-resolution using the eigenface representation. In pixel-domain super-resolution, the observations are low-resolution images that are related to a high-resolution image with a linear mapping. By ordering images lexicographically, such a relation can be written in matrix-vector notation as follows:

$$\mathbf{y}^{(i)} = \mathbf{H}^{(i)}\mathbf{x} + \mathbf{n}^{(i)}, \qquad \text{for } \; i = 1, \ldots, M \tag{32}$$

where $\mathbf{x}$ is the unknown high-resolution image, $\mathbf{y}^{(i)}$ is the ith low-resolution image observation, $\mathbf{H}^{(i)}$ is a linear operator that incorporates the motion, blurring, and downsampling processes, $\mathbf{n}^{(i)}$ is the

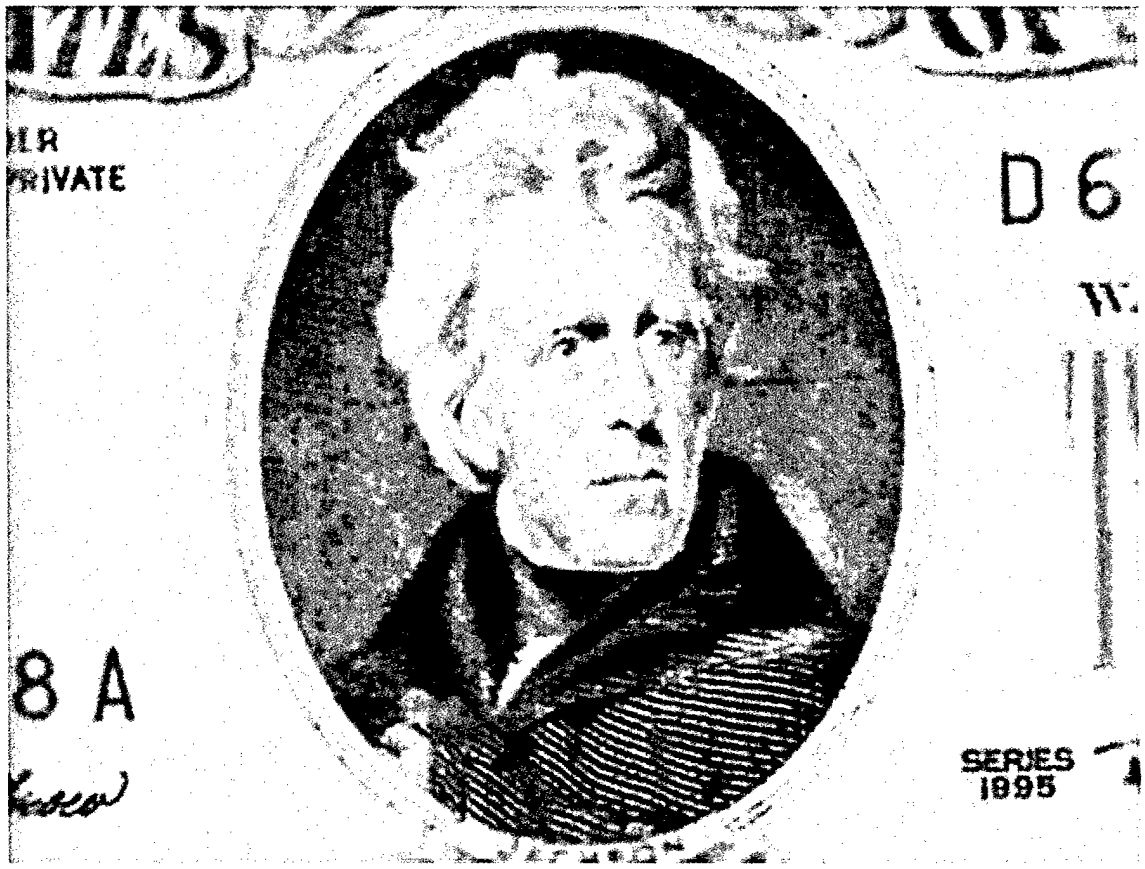


Figure 21: Reconstruction from the three-bit images shown in Figure 20.

noise vector, and M is the number of observations. Assuming that s is the downsampling factor ($0 < s < 1$), and that the high-resolution image is of dimension $N \times N$; $\mathbf{y}^{(i)}$, $\mathbf{H}^{(i)}$, $\mathbf{x}$, and $\mathbf{n}^{(i)}$ have dimensions $s^2N^2 \times 1$, $s^2N^2 \times N^2$, $N^2 \times 1$, and $N^2 \times 1$, respectively. Details of such modeling can be found in [21, 42], and we will not elaborate on it in this paper. The images $\mathbf{x}$ and $\mathbf{y}^{(i)}$ have two components that are in and orthogonal to the face space. Only the components that lie in the face space are necessary in recognition. We will now derive the observation model for the reconstruction of the components that lie in the face space. The formulation and reconstruction algorithm will not neglect the spatial-domain observation noise and the subspace representation error, which is initially orthogonal to the face space but then becomes effective during the imaging process. We start with writing the face space representation:

$$\mathbf{x} = \mathbf{\Phi}\mathbf{a} + \mathbf{e}_{\mathbf{x}}, \tag{33}$$

$$\mathbf{y}^{(i)} = \mathbf{\Psi}\hat{\mathbf{a}}^{(i)} + \mathbf{e}_{\mathbf{y}}^{(i)}, \qquad \text{for } \; i = 1, \ldots, M \tag{34}$$

where $\mathbf{\Phi}$ and $\mathbf{\Psi}$ are $N^2 \times L$ and $s^2N^2 \times L$ matrices that contain eigenfaces in their columns, $\hat{\mathbf{a}}^{(i)}$ is the $L \times 1$ feature vector that is associated with the ith observation, and $\mathbf{e}_{\mathbf{x}}$ and $\mathbf{e}_{\mathbf{y}}^{(i)}$ are the $N^2 \times 1$ and $s^2N^2 \times 1$ representation error vectors. Note that we have two different eigenvector bases, $\mathbf{\Phi}$ and $\mathbf{\Psi}$, corresponding to high and low resolution face images, respectively. (If we had included an upsampling matrix in $\mathbf{H}^{(i)}$, then we could use the same basis matrix.)

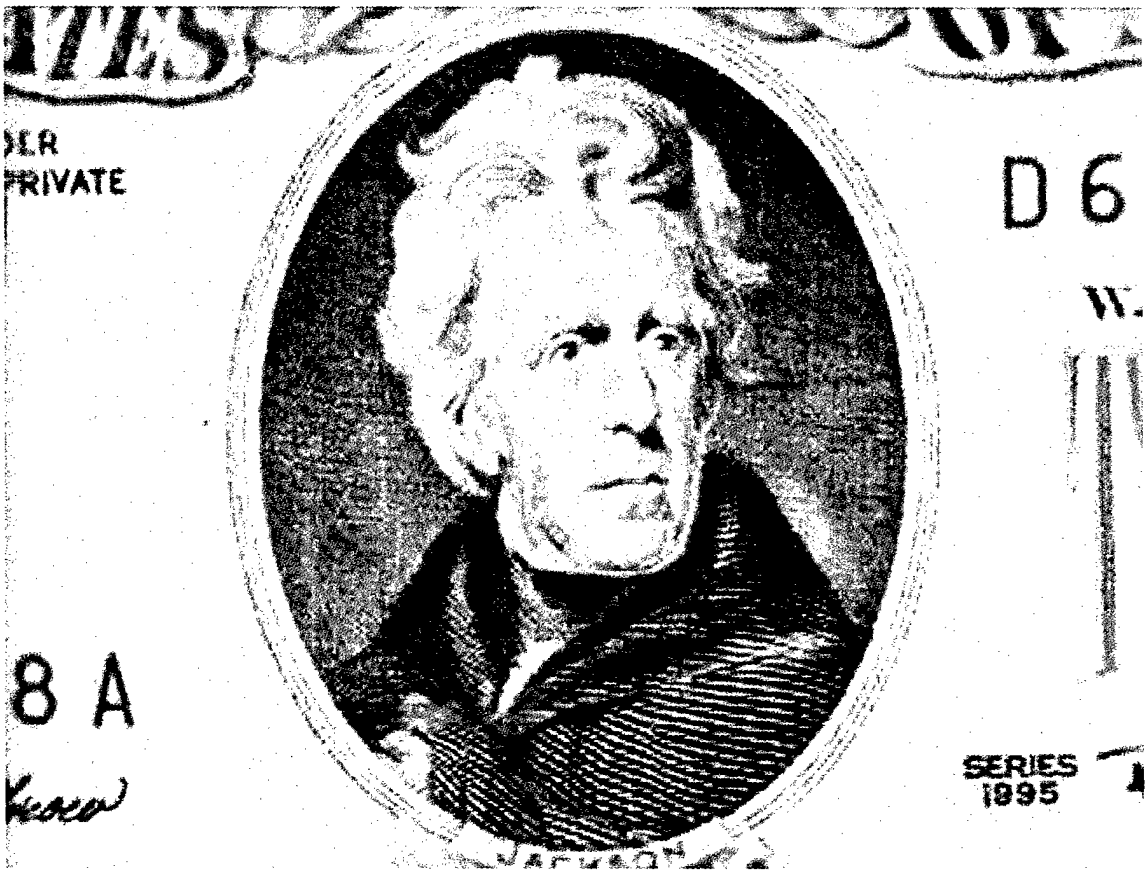


Figure 22: Reconstruction from eight compressed four-bit images.

We substitute Equations 33 and 34 into Equation 32 to obtain

$$\mathbf{\Psi}\hat{\mathbf{a}}^{(i)} + \mathbf{e}_{\mathbf{y}}^{(i)} = \mathbf{H}^{(i)}\mathbf{\Phi}\mathbf{a} + \mathbf{H}^{(i)}\mathbf{e}_{\mathbf{x}} + \mathbf{n}^{(i)}. \tag{35}$$

Now, we will project Equation 35 into the lower-dimensional face space using the fact that the representations errors $\mathbf{e}_{\mathbf{y}}^{(i)}$ are orthogonal to the face space $\mathbf{\Psi}$. Since

$$\mathbf{\Psi}^T\mathbf{e}_{\mathbf{y}}^{(i)} = \mathbf{0}, \qquad \text{for } \ i = 1, \ldots, M, \tag{36}$$

and

$$\mathbf{\Psi}^T\mathbf{\Psi} = \mathbf{I}, \tag{37}$$

by multiplying both sides of Equation 35 by $\mathbf{\Psi}^T$ on the left, we obtain:

$$\hat{\mathbf{a}}^{(i)} = \mathbf{\Psi}^T\mathbf{H}^{(i)}\mathbf{\Phi}\mathbf{a} + \mathbf{\Psi}^T\mathbf{H}^{(i)}\mathbf{e}_{\mathbf{x}} + \mathbf{\Psi}^T\mathbf{n}^{(i)}. \tag{38}$$

This is the observation equation that is analogous to Equation 32. It gives the relation between the unknown "true" feature vector $\mathbf{a}$ and the observed "inaccurate" feature vectors $\hat{\mathbf{a}}^{(i)}$. In the traditional way of applying super-resolution, the unknown high-resolution image $\mathbf{x}$ in Equation 32 is reconstructed from the low-resolution observations $\mathbf{y}^{(i)}$. Then, the reconstructed $\mathbf{x}$ is fed into a face recognition system. (See Figure 28.) For eigenface-based face recognition systems, a better way is to directly reconstruct the low-dimensional feature vector. Using the relation provided in

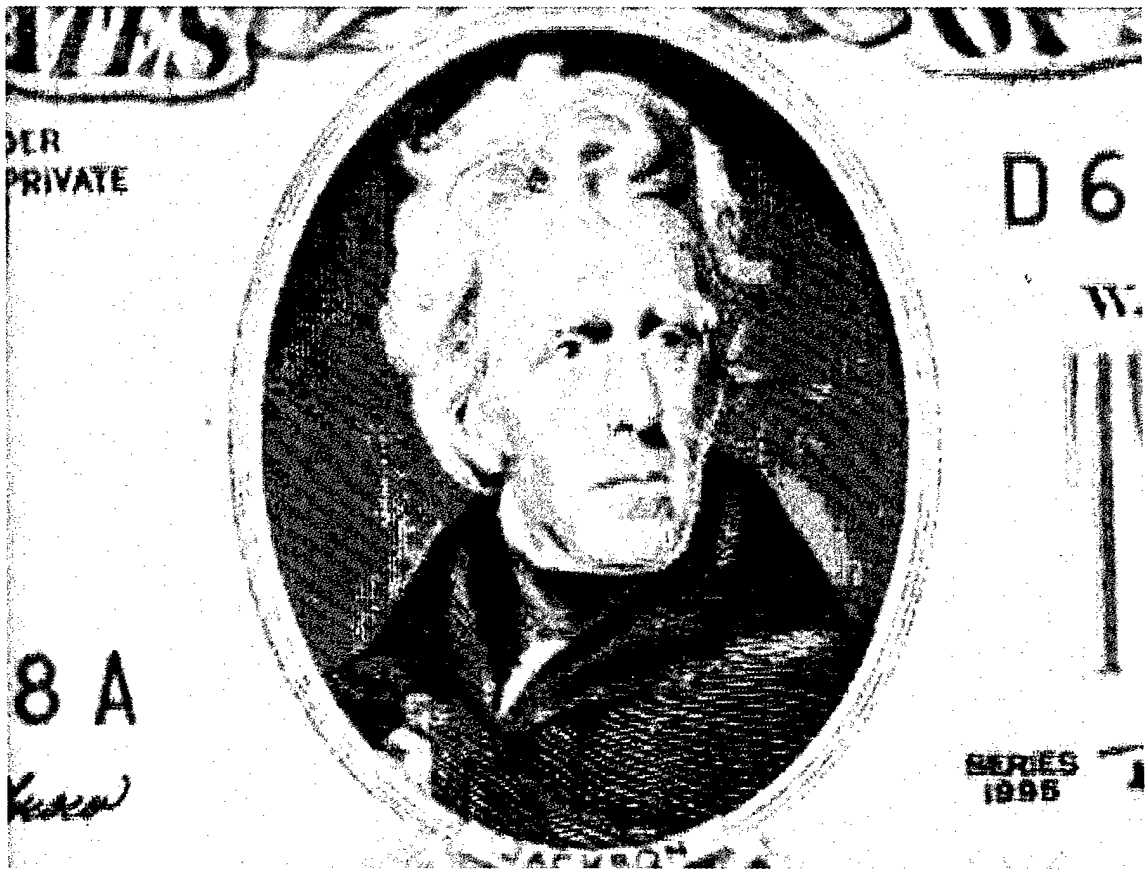


Figure 23: Reconstruction from eight compressed three-bit images.

Equation 38, accurate feature vectors of a face image can be obtained from its inaccurate feature vector observations. This is illustrated in Figure 29. The face observations $\mathbf{y}^{(i)}$ are first projected to the face space, and the computationally intensive super-resolution reconstruction is performed in the low-dimensional face subspace instead of the spatial domain. While we are reconstructing the feature vectors in the low-dimensional subspace, we will substitute face specific information in the form of statistics of the prior distributions of the feature vectors and distributions of the noise processes. Since all of these information is transformed to the low-dimensional face space, the computational complexity is kept low with little or no sacrifice of the performance. Also, using model-based information in regularization helps to obtain more robust results when compared to the traditional super-resolution algorithms.

4.5.3 Reconstruction algorithm

In this section we present a reconstruction algorithm to solve Equation 38 based on Bayesian estimation. The algorithm handles the observation noise and subspace representation error in the low-dimensional face subspace. The maximum *a posteriori* probability (MAP) estimator $\tilde{\mathbf{a}}$ is the argument that maximizes the product of the conditional probability $p(\hat{\mathbf{a}}^{(1)}, \cdots, \hat{\mathbf{a}}^{(M)}|\mathbf{a})$ and the

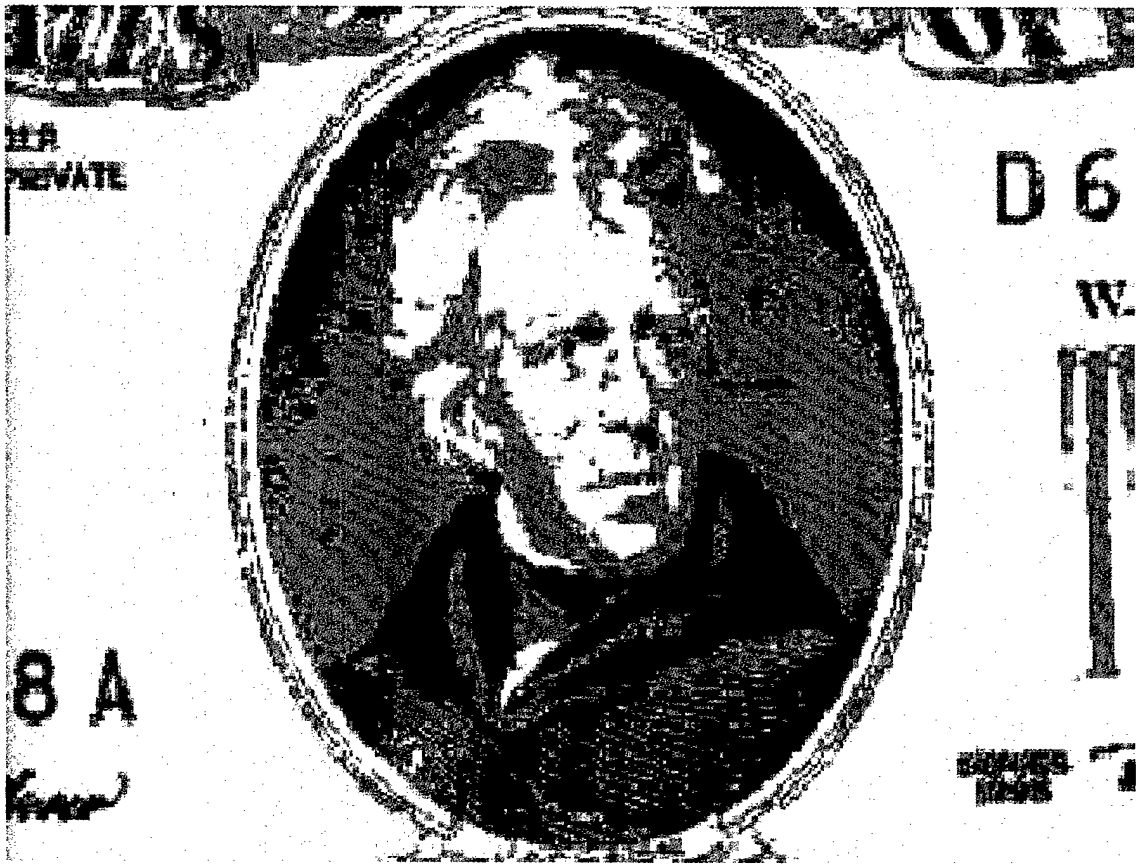


Figure 24: Low resolution two-bit MONEY image (one of eight).

prior probability $p(\mathbf{a})$:

$$\tilde{\mathbf{a}} = \arg\max_{\mathbf{a}} \left\{ p(\hat{\mathbf{a}}^{(1)}, \cdots, \hat{\mathbf{a}}^{(M)} | \mathbf{a}) p(\mathbf{a}) \right\}. \tag{39}$$

We now need to model the statistics $p(\hat{\mathbf{a}}^{(1)}, \cdots, \hat{\mathbf{a}}^{(M)} | \mathbf{a})$ and $p(\mathbf{a})$. The prior probability $p(\mathbf{a})$ can simply be assumed as jointly Gaussian:

$$\begin{aligned} p(\mathbf{a}) &= \tfrac{1}{Z} \exp\left(-\left(\mathbf{a} - \mu_{\mathbf{a}}\right)^T \mathbf{\Lambda}^{-1} \left(\mathbf{a} - \mu_{\mathbf{a}}\right) \right) \\ &\triangleq \mathcal{N}\left(\mu_{\mathbf{a}}, \mathbf{\Lambda}\right), \end{aligned} \tag{40}$$

where $\mathbf{\Lambda}$ is the $L \times L$ covariance matrix, $\mu_{\mathbf{a}}$ is the $L \times 1$ mean of $\mathbf{a}$, and Z is a normalization constant. We also introduced the notation $\mathcal{N}(\cdot, \cdot)$ for normal distribution to simplify the notation for the rest of the paper.

In order to find $p(\hat{\mathbf{a}}^{(1)}, \cdots, \hat{\mathbf{a}}^{(M)} | \mathbf{a})$, we first model the noise process in the spatial domain, and then derive the statistics in face space. We define a total noise term $\mathbf{v}^{(i)}$ that consists of the noises resulting from the subspace representation error $\mathbf{e_x}$ and the observation noise $\mathbf{n}^{(i)}$ in the spatial domain:

$$\mathbf{v}^{(i)} \triangleq \mathbf{H}^{(i)} \mathbf{e_x} + \mathbf{n}^{(i)}. \tag{41}$$

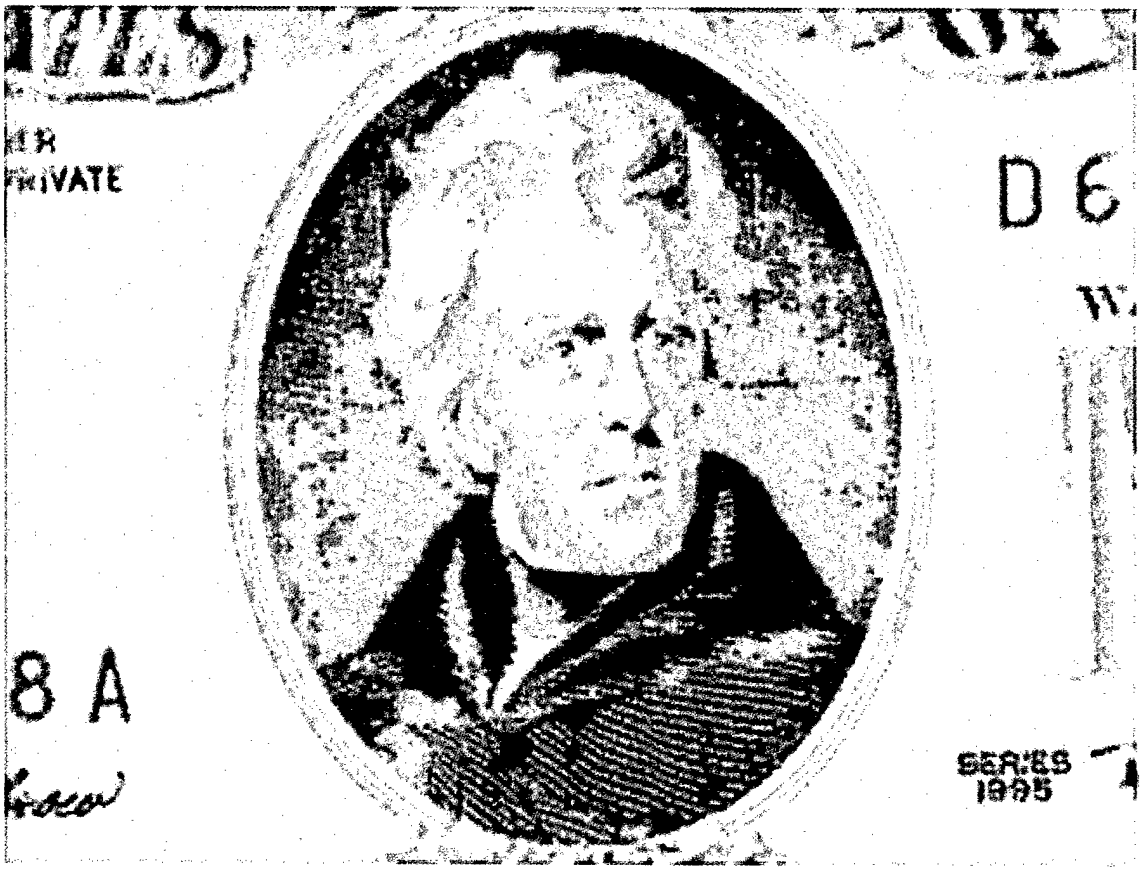


Figure 25: Reconstruction from the two-bit images shown in Figure 24.

Using this definition, we rewrite Equation 38 for convenience:

$$\hat{\mathbf{a}}^{(i)} = \mathbf{\Psi}^T \mathbf{H}^{(i)} \mathbf{\Phi} \mathbf{a} + \mathbf{\Psi}^T \mathbf{v}^{(i)}. \tag{42}$$

The reason we defined $\mathbf{H}^{(i)}\mathbf{e_x} + \mathbf{n}^{(i)}$ as the total noise term instead of its projection onto the face subspace is because of the modeling convenience in the spatial domain. It has been demonstrated that modeling the noise (resulting from the imaging system and the estimation of $\mathbf{H}^{(i)}$) in the spatial domain as independent identically distributed (IID) Gaussian random vectors is a good assumption [21]. We further assume that the covariance matrix of these random vectors is diagonal so that the statistical parameters can be estimated easily even with the limited training data. Using these assumptions, it is easy to find the distribution of $\mathbf{\Psi}^T \mathbf{v}^{(i)}$ in the face space, as will be shown shortly.

Defining $\mathbf{K}$ as the $s^2N^2 \times s^2N^2$ positive definite diagonal covariance matrix and $\mu_{\mathbf{v}}^{(i)}$ as the $s^2N^2 \times 1$ mean of $\mathbf{v}^{(i)}$, we can write the probability distribution of $\mathbf{v}^{(i)}$ as:

$$p(\mathbf{v}^{(i)}) = \mathcal{N}\left(\mu_{\mathbf{v}}^{(i)}, \mathbf{K}\right). \tag{43}$$

Now, we need to derive the distribution of the projected noise, $p(\mathbf{\Psi}^T \mathbf{v}^{(i)})$, in order to get the conditional PDF $p(\hat{\mathbf{a}}^{(1)}, \cdots, \hat{\mathbf{a}}^{(M)} | \mathbf{a})$. From the analysis of functions of multi-variate random variables,

Figure 26: Low resolution one-bit MONEY image (one of eight).

it follows that $p(\mathbf{\Psi}^T\mathbf{v}^{(i)})$ is also jointly Gaussian since $\mathbf{\Psi}^T\mathbf{\Psi}$ is nonsingular (by construction). As a result, we have:

$$p(\psi^T\mathbf{v}^{(i)}) = \mathcal{N}\left(\mathbf{\Psi}^T\mu_{\mathbf{v}}^{(i)}, \mathbf{Q}\right), \tag{44}$$

where $\mathbf{\Psi}^T\mu_{\mathbf{v}}^{(i)}$ is the new mean, and $\mathbf{Q}$ is the new covariance matrix computed by $\mathbf{Q} = \mathbf{\Psi}^T\mathbf{K}\mathbf{\Psi}$. The covariance matrix $\mathbf{Q}$ has dimension of $L \times L$ while $\mathbf{K}$ is of dimension $s^2N^2 \times s^2N^2$. Using Equations 42 and 44, we find the conditional PDF $p(\hat{\mathbf{a}}^{(i)}|\mathbf{a})$:

$$p(\hat{\mathbf{a}}^{(i)}|\mathbf{a}) = \mathcal{N}\left(\mathbf{\Psi}^T\mathbf{H}^{(i)}\mathbf{\Phi}\mathbf{a} + \mathbf{\Psi}^T\mu_{\mathbf{v}}^{(i)}, \mathbf{Q}\right). \tag{45}$$

Since we assumed that $\mathbf{v}^{(i)}$'s are IID, it follows that the probability density function $p(\hat{\mathbf{a}}^{(1)}, \cdots, \hat{\mathbf{a}}^{(M)}|\mathbf{a})$ is the product of $p(\hat{\mathbf{a}}^{(i)}|\mathbf{a})$ for $i = 1, \cdots, M$. Defining $\xi^{(i)} \triangleq \hat{\mathbf{a}}^{(i)} - \mathbf{\Psi}^T\mathbf{H}^{(i)}\mathbf{\Phi}\mathbf{a} - \mathbf{\Psi}^T\mu_{\mathbf{v}}^{(i)}$, we write:

$$p(\hat{\mathbf{a}}^{(1)}, \cdots, \hat{\mathbf{a}}^{(M)}|\mathbf{a}) = \frac{1}{Z}\exp\left(-\sum_{i=1}^{M}\xi^{(i)T}\mathbf{Q}^{-1}\xi^{(i)}\right), \tag{46}$$

where Z is a normalization constant.

Substituting the conditional and prior PDFs given in Equations 40 and 46 into Equation 39 , we

Figure 27: Reconstruction from the one-bit images shown in Figure 26.

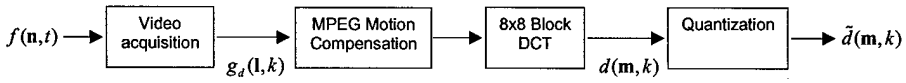


Figure 28: Super-resolution applied as a preprocessing block to face recognition.

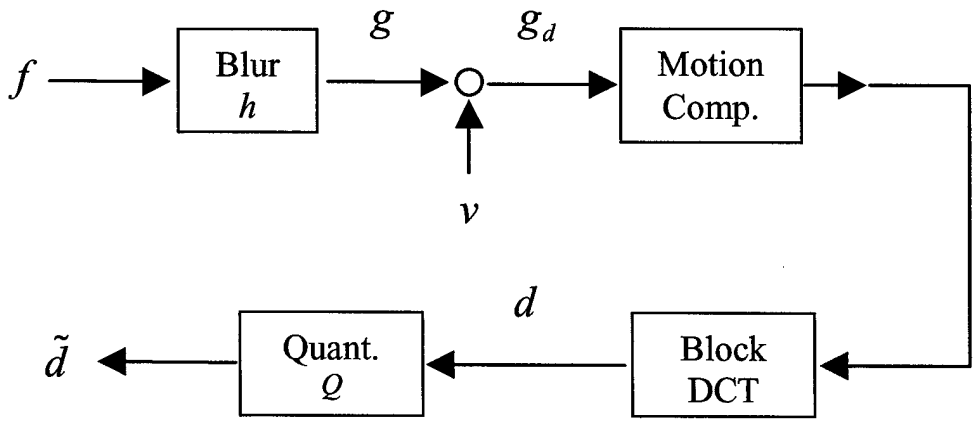


Figure 29: Super-resolution embedded into eigenface-based face recognition.

obtain the MAP estimate $\tilde{\mathbf{a}}$ as follows:

$$\tilde{\mathbf{a}} = \arg\min_{\mathbf{a}} \left\{ \sum_{i=1}^{M} \left[\xi^{(i)^T} \mathbf{Q}^{-1} \xi^{(i)} \right] + (\mathbf{a} - \mu_{\mathbf{a}})^T \mathbf{\Lambda}^{-1} (\mathbf{a} - \mu_{\mathbf{a}}) \right\}. \tag{47}$$

This solution can be obtained easily using the iterative steepest descent algorithm.

In the reconstruction, everything but $\mathbf{a}$, $\mathbf{a}^{(i)}$, and $\mathbf{H}^{(i)}$ is known and can be computed in advance. For a specific observation sequence $\mathbf{y}^{(i)}$, the feature vectors $\mathbf{a}^{(i)}$ and the blur mappings $\mathbf{H}^{(i)}$ are computed, and the true feature vector $\mathbf{a}$ is estimated iteratively. Each iteration in such algorithm requires a number of operations that is directly proportional to the size of the vector to be reconstructed. If super-resolution is applied as a preprocessing block, then the number of operations will be $O(N^2)$, N^2 being the number of pixels in the face image. For the face-space reconstruction algorithm, the number of operations reduces to $O(L)$. Therefore, face-space super-resolution provides an efficiency gain proportional to N^2/L in computation over pixel-domain super-resolution. (Typically face images are of size 60×60 and a face dimension of $L = 50$ is satisfactory, in which case the face-space super-resolution is approximately 72 times faster than the pixel-domain super-resolution.)

4.5.4 Experimental results

We performed a set of experiments to demonstrate the efficacy of the proposed method. We investigated the effect of the face space dimension, and sensitivity to noise and motion estimation errors.

In these experiments, we used face images from the Yale face databases A and B [24], Harvard Robotics Laboratory database [31], and AR database [44]. The images are downsampled to have a size of 40×40, and aligned according to the manually located eye and mouth locations. We selected 134 images as training data and 50 images as test data. We applied KLT to those 134 images and chose the first 60 eigenvectors having the largest eigenvalues to form the face subspace. (These 60 eigenvectors form the columns of the matrix $\mathbf{\Phi}$.) We also downsampled the training images by four to obtain 10×10 images, applied the KLT to those images, and chose the first 60 of them to construct the eigenface space $\mathbf{\Psi}$.

The test images were jittered by a random amount to simulate motion, blurred, and downsampled by a factor of four to produce multiple low-resolution images for each subject. The motion vectors were saved for use in synthetic video experiments. For blurring, the images were convolved with a point spread function (PSF), which was set to a 5×5 normalized Gaussian kernel with zero mean and a standard deviation of one pixel.

From the training image set, ($K = 134$), we estimate the statistics of $\mathbf{a}$ and $\mathbf{v}^{(i)}$. The unbiased estimates for the mean and covariance matrix of $\mathbf{a}$ are simply obtained from the sample mean and variances.

One of the frames for each video sequence is chosen as the reference frame, bilinearly interpolated by four, and projected onto the face space $\mathbf{\Phi}$ to obtain the initial estimate for the true feature vector. It is then updated using the steepest descent technique. The mapping $\mathbf{H}^{(i)}$ is computed from the known motion vectors and PSF, and 16 low-resolution images are used in the reconstruction.

We also wanted to compare the results of this eigenface-domain super-resolution algorithm with those of traditional pixel-domain super-resolution. We applied the pixel-domain super-resolution algorithm given in [42] to the low-resolution video sequences again using the same 16 low-resolution images and setting the number iterations to seven. After the high-resolution images are reconstructed, they are projected onto the face space $\mathbf{\Phi}$ to obtain the feature vectors.

The feature vectors obtained from these algorithms are compared with the true feature vectors (which are computed using the 40×40 original high-resolution images). Figure 30 shows the results for the normalized distance between the true feature vector $\mathbf{a}$ and the estimated feature vector $\tilde{\mathbf{a}}$. Figure 31 provides an example from the face database. It is seen that eigenface-domain super-resolution performs as well as the pixel-domain super-resolution at a lower computational complexity.

4.5.5 Conclusions

In this paper, we propose to apply super-resolution after dimensionality reduction in a face recognition system. In this way, only the necessary information for recognition is reconstructed. We have also shown how to incorporate the model-based information into the face-space reconstruction algorithm. This helps to obtain more robust results when compared to the traditional super-resolution algorithms. We investigate the effects the effects of the feature vector length (*i.e.*, dimension of the face space), noise, and motion estimation error on the performance. The detailed results will be pr

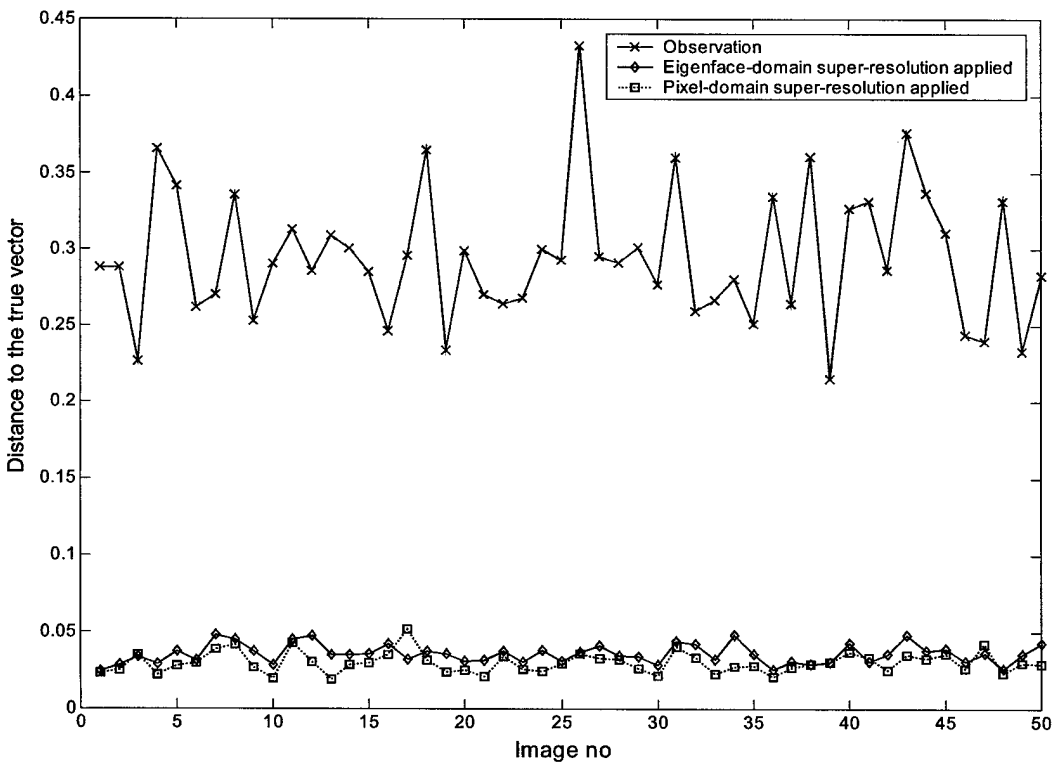


Figure 30: Error in feature vector computation.

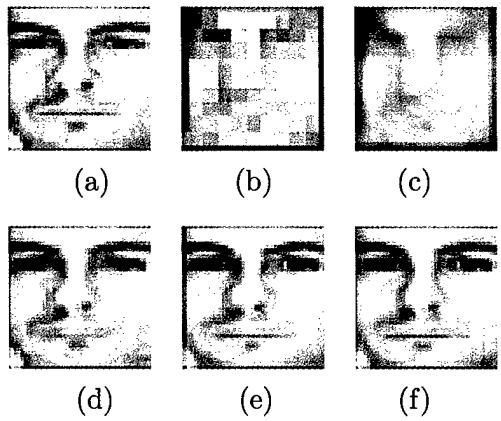


Figure 31: (a) Original 40×40 image. (b) 10×10 low-resolution observation is interpolated using nearest neighbor interpolation. (c) 10×10 low-resolution observation is interpolated using bilinear interpolation. (d) Pixel-domain super-resolution applied. (e) The result of pixel-domain super-resolution reconstruction is projected into the face subspace. (f) Representation of the feature vector reconstructed using the eigenface-domain super-resolution in the face subspace.

4.6 Effects of Camera Response Function and Illumination Changes in Multi-Frame Image Reconstruction

An image is a two-dimensional projection of a real-valued scene $f(t; \lambda; x, y, z)$ that is a function of time t, spectral wavelength λ, and space (x, y, z). During an imaging process, the quantity $f(t; \lambda; x, y, z)$ is degraded in several ways. The degradation may be on the domain $(t; \lambda; x, y, z)$ as well as on the range f. 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 does. (Image fusion is sometimes referred to as displaying images on top of each other to help decision making. For instance, in surveillance systems, visible and infra-red (IR) data are used to create a composite where people only visible in the IR imagery are combined with the context of the terrain from visible image. Here, we refer to image fusion as reconstruction of original scene.) With image fusion, we can achieve four types of improvement: temporal, spectral, space, and range.

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 observed incorrectly. This is known as motion aliasing. Also, due to the nonzero exposure, fast motion causes the so-called motion blur, integration of light on the sensor array over time. It is possible to catch events that are faster than the frame rate when the motion can be modeled [51] and/or when there are multiple cameras capturing the scene with a time offset or at a different frame rate [56].

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 provide sufficient visual quality for most of the time. However, in multi-spectral and hyper-spectral imagery, high spatial and spectral resolution is desirable, and various algorithms fusing images of different spectral and spatial resolutions have been proposed [25, 26].

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

improvement in space. The improvement can be three-dimensional (3D) reconstruction when the observations and the scene model provide sufficient information about the 3D structure. In some applications, two-dimensional improvement is sought. For instance, super-resolution reconstruction improves spatial resolution by sub-pixel registration of images [42, 34, 16, 53, 21]. In mosaic construction, an image of larger spatial extent is constructed [61]. Color filter array interpolation is another example image fusion where spatial resolution of the color channels are improved using the correlation between red, green, and blue spectral data [27].

D. Range Improvement: Although a real scene has in general a wide dynamic range, imaging devices are limited in dynamic range. During an imaging process, the dynamic range of a scene is degraded by noise, dynamic range compression, and quantization. Multi-frame filtering algorithms have been widely used to get rid of the noise. Recent works by Mann, Robertson *et al.*, and Candocia improve the dynamic range (range extent) of the images [43, 48, 10, 11].

In this project, we are proposing an image fusion algorithm that improves resolution and extent in both range and spatial domain. Although there is a large amount of work done aiming to improve spatial resolution, none of these super-resolution algorithms consider improving range. However, the observations may have information diversity in range due to changes of illumination in the scene or imaging device adjustments such as exposure time, gain, white balance; and this can be used to produce composite images of higher dynamic range (range extent) and tonal fidelity (range resolution) in addition to the higher spatial resolution and extent.

Next section presents an observation model that establishes the connection between a scene and multiple observations of that scene that are limited in spatial and range resolution/extent. Based on this model, an image fusion algorithm that achieves improvement in both range and spatial domain is proposed in Section 3. Section 4 addresses some of the implementation issues, and provides preliminary results.

4.6.1 Imaging model

In this section, we will extend a typical imaging model used in super-resolution algorithms to include dynamic range compression and quantization operations. 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), blur (because of the nonzero sensor aperture time, the nonzero physical dimensions of the individual sensor elements, the degree of focus, etc.), and sampling with a low-resolution grid [42]. According to this model, the mapping from a high-resolution image to a low spatial-resolution image is expressed as a weighted sum of the high-resolution image pixels, where the weights are the values of a space-invariant point-spread function (PSF) at the corresponding pixel locations. The center of the PSF depends upon the motion between the high-resolution image and the low-resolution images. Motion vectors from each low spatial-resolution image to the high-resolution image determine how each pixel is mapped. The normalized PSF that characterizes the camera is centered at that location, and from it the weights on the high-resolution image grid are found. Defining $h_i(m_1, m_2; n_1, n_2)$ as this mapping, we can model the image capture process by:

$$z_i(m_1, m_2) = \sum_{n_1, n_2} h_i(m_1, m_2; n_1, n_2) f(n_1, n_2), \tag{48}$$

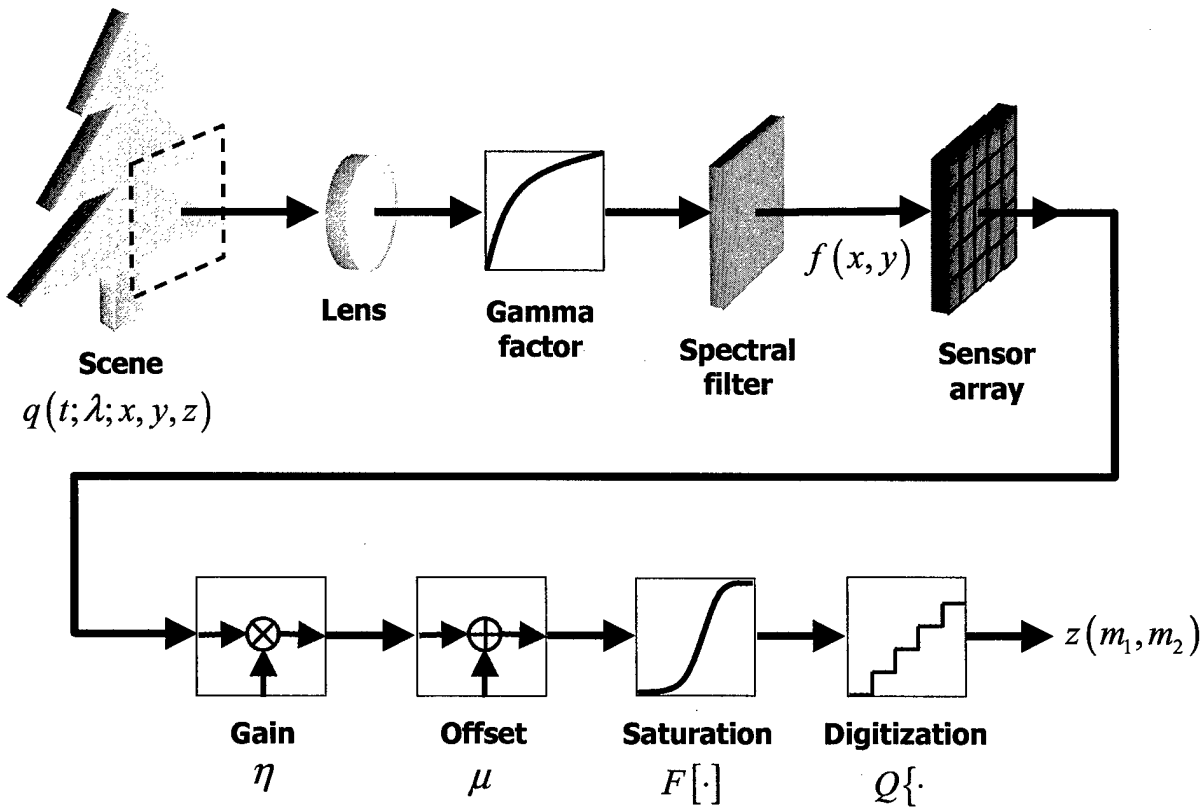


Figure 32: Imaging model.

where (n_1, n_2) and (m_1, m_2) are discrete coordinates of high- and low- resolution images, respectively, and i is the observation number. This model assumes that the observations are captured under the same illumination condition and camera settings. However, this assumption is not always valid. It is likely that measurements of the same scene are obtained with different camera settings (such as exposure time, gamma factor, offset, etc.) in addition to potential illumination changes in the scene itself.

We now extend the model formulated in (58) to include processes affecting the range of the observation. The extended model is given in Figure 39. In addition to the previous model, this model also includes the effects of exposure timer, dynamic range compressor, and quantizer. Exposure timer determines the amount of light falling on the sensor array. In order to see dark regions in a scene, exposure time should be set long. With a long exposure time, light regions in a scene get saturated. (When we have multiple images of the same scene captured with different exposure times, we can register those images and obtain a composite image of larger dynamic range [43, 48, 10, 11].) After dynamic range is compressed, pixel intensities are digitized. Usually, eight bits are used for each pixel.

Including the effects acting on range, we can extend the equation given in (58) as follows:

$$z_i(m_1, m_2) = Q\left\{F\left[\eta_i\left(\sum_{n_1,n_2} h_i(m_1, m_2; n_1, n_2) f(n_1, n_2)\right) + \mu_i\right]\right\}, \tag{49}$$

where $F[\cdot]$ is the dynamic range compression function, $Q\{\cdot\}$ is the quantizer, η_i and μ_i are the illu-

mination gain and offset, respectively. In this model, the illumination gain and offset are considered space-invariant. This is a valid assumption unless there is non-uniform illumination change within the scene. A typical dynamic range compression function is depicted in Figure 40. It is linear (or log-linear) at midtones, and becomes flat at low and high intensities. The quantizer is assumed to have uniform step sizes; therefore, the joint effect of dynamic range compression function and quantizer is a quantizer of non-uniform step sizes. The quantization noise is larger at low and high pixel intensities compared to the midtones.

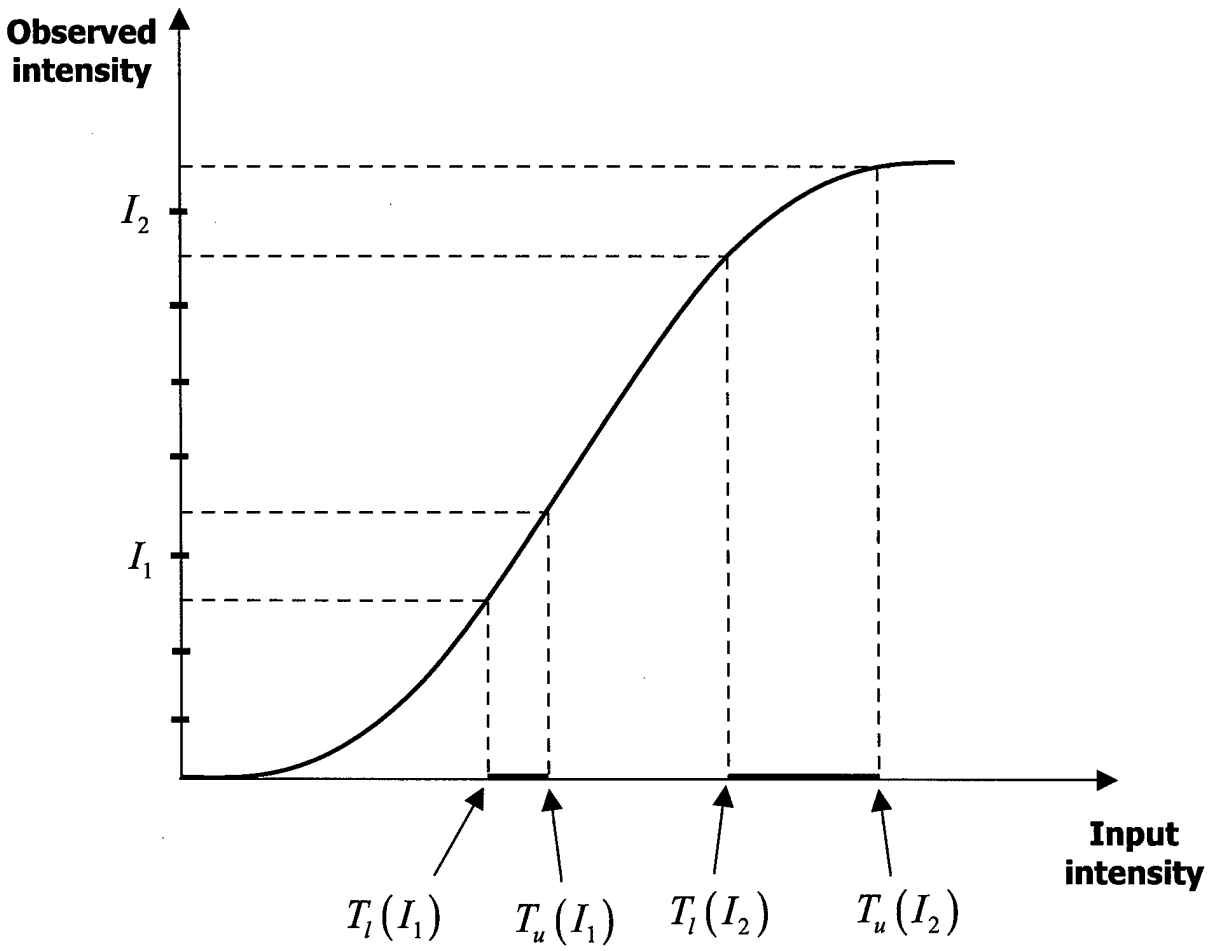


Figure 33: Bounds of the input pixel intensity can be determined from the observed pixel intensity. The bounds are looser towards the extremes because of larger quantization step sizes.

4.6.2 Joint Spatial and Range Improvement

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

As formulated in Equation (59), dynamic range compression and quantization introduces quanti-

zation error in the pixel intensities of the observations. The quantization error is error small at the midtones, but large towards the ends of the dynamic range. Let $\delta_i(m_1, m_2)$ be the quantization error at pixel (m_1, m_2) of ith observation, then we can rewrite Equation (59) as:

$$z_i(m_1, m_2) = \eta_i \left(\sum_{n_1, n_2} h_i(m_1, m_2; n_1, n_2) f(n_1, n_2) \right) + \mu_i + \delta_i(m_1, m_2). \tag{50}$$

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 dynamic range compression response of the camera is known. This is illustrated in Figure 40. The lower and upper bounds of the quantization error are closer to each other for midtone pixel intensities. The measurements towards the ends of the dynamic range has wider bounds.

We now define constraint sets using the bounds of range data, and employ a projections onto convex sets (POCS) algorithm. Let $T_l(q)$ and $T_u(q)$ be the lower and upper bounds of quantization for a measured pixel intensity q, then we can write the constraint set on any measurement $z_i(m_1, m_2)$ as follows:

$$C\left[z_i(m_1, m_2)\right] = \left\{ f(n_1, n_2) : T_l\left(z_i(m_1, m_2)\right) \leq |r_i(m_1, m_2)| \leq T_u\left(z_i(m_1, m_2)\right) \right\}, \tag{51}$$

where $x(n_1, n_2)$ is the initial estimate of the original scene $f(n_1, n_2)$, and $r_i(m_1, m_2)$ is the residual between the observed pixel intensity and the intensity computed from the initial estimate:

$$r_i(m_1, m_2) = z_i(m_1, m_2) - \left[\eta_i \left(\sum_{n_1, n_2} h_i(m_1, m_2; n_1, n_2) x(n_1, n_2) \right) + \mu_i \right]. \tag{52}$$

Set-theoretic reconstruction techniques produce solutions that are consistent with the information arising from observed data or prior knowledge about the solution. Each piece of information is associated with a constraint set in the solution space, and the intersection of these sets represents the space of acceptable solutions [19]. When there are multiple observations, the intersection of the constraint sets may result in finer range resolution and dynamic range. This is illustrated in Figure 41.

For each pixel in the observations, we can define a constraint set, and project an initial estimate onto these constraint set to improve its resolution and extent in range and spatial domain. After some simple algebra, the projection operator onto a constraint set $C\left[z_i(m_1, m_2)\right]$ can be found as:

$$P_{C[z_i(m_1,m_2)]}\left[x(n_1, n_2)\right] = x(n_1, n_2) + \left\{ \begin{array}{c} \frac{1}{\eta_i} \left(\frac{(r_i(m_1,m_2) - T_u(z_i(m_1,m_2)))h_i(m_1,m_2;n_1,n_2)}{\sum_{n_1,n_2} |h_i(m_1,m_2;n_1,n_2)|^2} - \mu_i \right), r_i(m_1, m_2) > T_u\left(z_i(m_1, m_2)\right) \\ 0, T_l\left(z_i(m_1, m_2)\right) \leq |r_i(m_1, m_2)| \leq T_u\left(z_i(m_1, m_2)\right) \\ \frac{1}{\eta_i} \left(\frac{(r_i(m_1,m_2) - T_l(z_i(m_1,m_2)))h_i(m_1,m_2;n_1,n_2)}{\sum_{n_1,n_2} |h_i(m_1,m_2;n_1,n_2)|^2} - \mu_i \right), r_i(m_1, m_2) < T_l\left(z_i(m_1, m_2)\right) \end{array} \right\}. \tag{53}$$

In [42], a projections onto convex sets (POCS) algorithm is presented. In that algorithm, bounds of the constraint sets are set to a fixed number, which is chosen heuristically. In this paper, we

show that those bounds should actually be a function of observed pixel intensity, and should be chosen using the camera response function.

In the next section, we give the details of estimating registration parameters, obtaining initial estimate, and the complete algorithm.

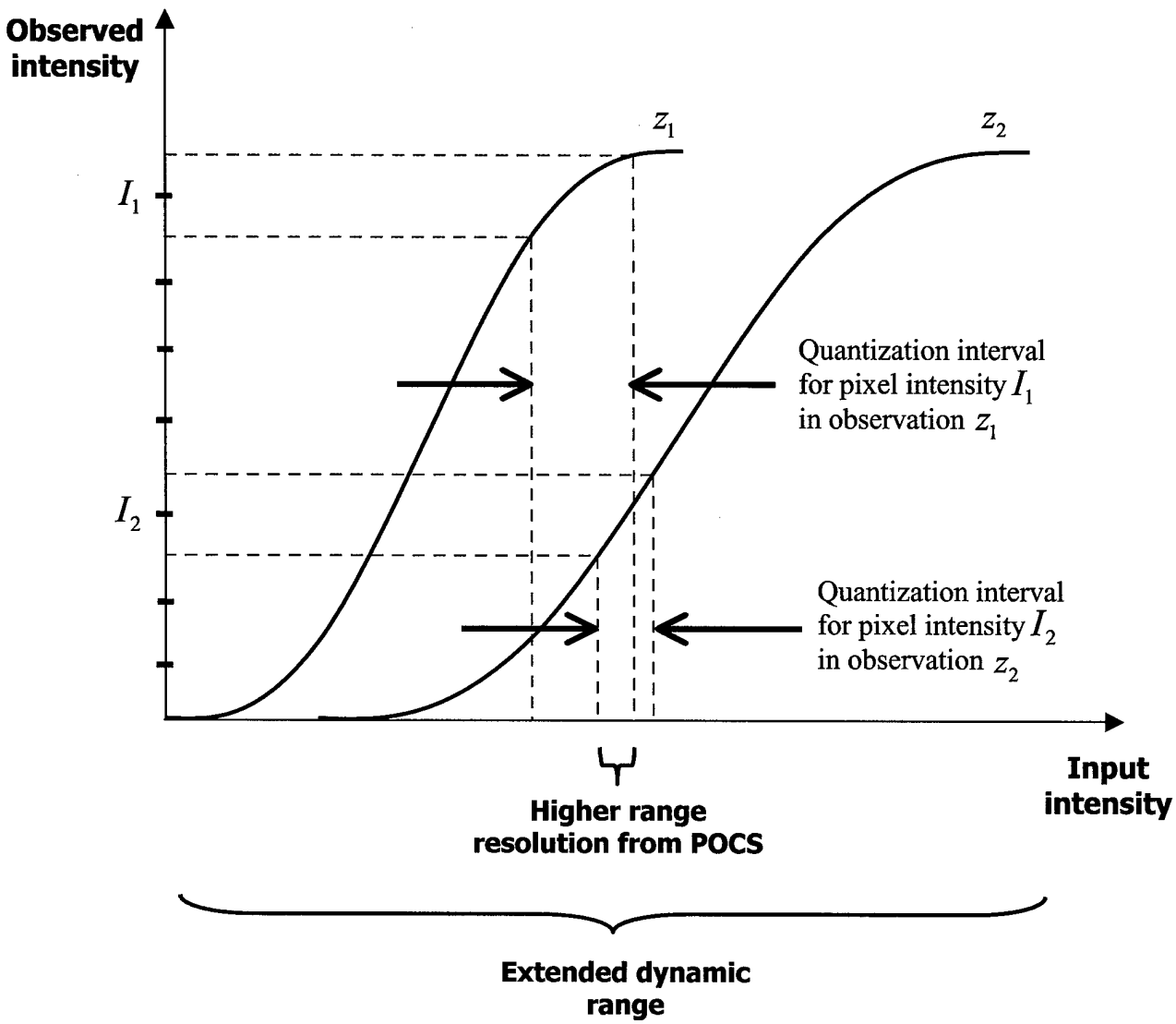


Figure 34: Multiple observations lead to higher resolution and wider dynamic range.

4.6.3 Estimating camera response function, motion and illumination parameters

There are various ways of estimating camera response function, and motion and illumination parameters [43, 48, 10, 11]. Although all parameters can be estimated jointly, it is more reliable to estimate camera response function as a preprocessing step, and then use it in determination of motion and illumination parameters. Our initial approach is to linearize the camera response function and determine its parameters [10]. For the motion and illumination estimation, we use a

parametric motion model, and determine its parameters jointly with the illumination parameters. We now outline this approach for affine motion model; it can be extended to other models easily. Given the affine model parameters $[a_1\ a_2\ a_3\ a_4\ a_6\ a_6]$, the motion vectors are:

$$\begin{array}{l} u(x,y) = a_1x + a_2y + a_3 \\ v(x,y) = a_4x + a_5y + a_6 \end{array}. \tag{54}$$

Incorporating these equations into the optic flow equation with illumination model, we get:

$$\eta I_1(x,y) + \mu = I_2(x + a_1x + a_2y + a_3, y + a_4x + a_5y + a_6), \tag{55}$$

where $I_1(x,y)$ is the image intensity at pixel (x,y) in the first frame, $I_2(x+a_1x+a_2y+a_3, y+a_4x+a_5y+a_6)$ is the corresponding pixel intensity in the second frame, and η and μ are the illumination terms.

Applying Taylor series expansion to Equation (55) yields:

$$I_x(x,y)(a_1x + a_2y + a_3) + I_y(x,y)(a_4x + a_5y + a_6) + I_2(x,y) - \eta I_1(x,y) - \mu \cong 0. \tag{56}$$

We now define the cost function

$$\Psi = \sum_{x,y} \left\{ \begin{array}{l} I_x(x,y)(a_1x + a_2y + a_3) + I_y(x,y)(a_4x + a_5y + a_6) + \\ I_2(x,y) - \eta I_1(x,y) - \mu \end{array} \right\}^2, \tag{57}$$

and by taking the partial derivatives of Ψ with respect to affine and illumination parameters, and setting them to zero, we end up with a linear set of equations, which can be solved easily.

4.6.4 Experimental results

We provide the results of an experiment to demonstrate the usefulness of 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. The images from the video sequence are given in Figures 35 to 37. The image in 35 is captured with a short exposure time; the buildings outside the window can be seen clearly. On the other hand, the image in 37 is captured with a longer exposure time; the inside can be seen clearly, but outside the window cannot be seen due to saturation. We applied the proposed algorithm to these images. The reconstructed image is given in Figure 38. It has higher dynamic range than any of the observations does alone. Its dynamic range is from -22 to 227, which is scaled with respect to the first observation. (The pixel intensity range is given as a side bar.)

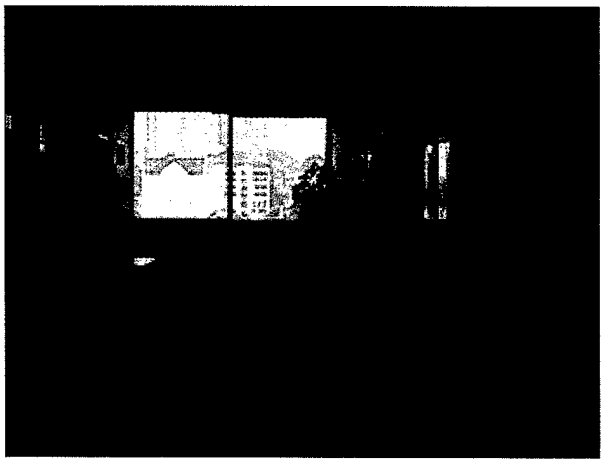


Figure 35: First observation.

Figure 36: Second observation.

Figure 37: Third observation.

Figure 38: Reconstructed image with higher spatial and range information.

4.7 Multi-Frame Information Fusion for Gray-Scale and Spatial Enhancement of Images

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 [51] and/or when there are multiple cameras that capture the scene with different time offsets or frame rates [56].

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 [26].

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 [42, 53].

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 [43], Robertson *et al.* [48], and Candocia [10] 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 [48] and Candocia [10] works, maximum likelihood and least squares estimation are used, respectively.

In this project, we propose a set-theoretic image fusion algorithm that improves resolution and extent in both the gray-scale (range) and spatial domains. Although considerable work has been done to improve the spatial resolution, none of these super-resolution algorithms 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.

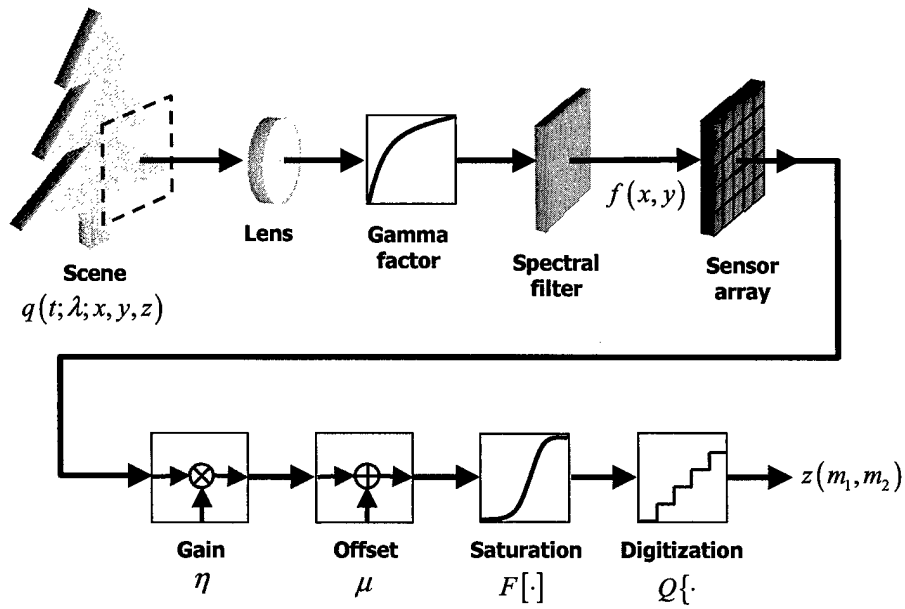


Figure 39: Imaging model includes effects acting on the gray-scale domain as well as on the spatial domain.

4.7.1 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 39. Light coming from the spectrally and spatially varying scene $q(t;\lambda;x,y,z)$ is passed through an optical system to 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

$$z_i(m_1, m_2) = \sum_{n_1, n_2} h_i(m_1, m_2; n_1, n_2) f(n_1, n_2), \tag{58}$$

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 [42, 53].

The model formulated in (58) 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 (58) as follows:

$$z_i(m_1, m_2) = Q\left\{F\left[\eta_i\left(\sum_{n_1, n_2} h_i(m_1, m_2; n_1, n_2) f(n_1, n_2)\right) + \mu_i\right]\right\}, \tag{59}$$

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. A typical dynamic range saturation function $F[\cdot]$ is depicted in Figure 40. Because of the nonlinearity of the saturation function, the quantization noise is larger at low and high pixel intensities than at the midrange.

4.7.2 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.

Constraint Sets from Amplitude Quantization

As formulated in (59), 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

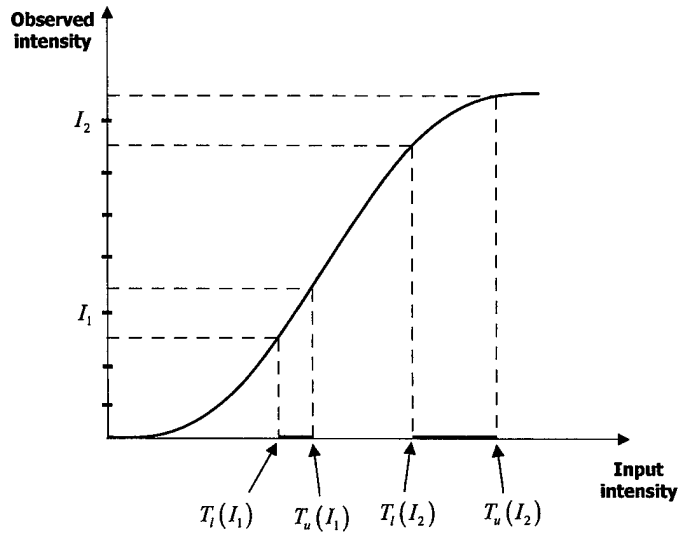


Figure 40: Pixel intensities are compressed and digitized observations of real-valued quantities. For each intensity level, there is a lower bound $T_l(\cdot)$ and an upper bound $T_u(\cdot)$ within which the input intensity lies.

pixel (m_1, m_2) of the ith observation. Then we can rewrite (59) as

$$z_i(m_1, m_2) = \eta_i \left(\sum_{n_1, n_2} h_i(m_1, m_2; n_1, n_2) f(n_1, n_2) \right) + \mu_i + \delta_i(m_1, m_2). \tag{60}$$

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 40, 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 [19]. 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 $f(n_1, n_2)$. Then we can write the constraint set on any observed pixel $z_i(m_1, m_2)$ as follows:

$$C\left[z_i(m_1, m_2)\right] = \left\{ \; f(n_1, n_2) : T_l\left(z_i(m_1, m_2)\right) < |\hat{z}_i(m_1, m_2)| < T_u\left(z_i(m_1, m_2)\right) \; \right\}, \tag{61}$$

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

$$\hat{z}_i(m_1, m_2) = \eta_i \left(\sum_{n_1, n_2} h_i(m_1, m_2; n_1, n_2) f(n_1, n_2) \right) + \mu_i. \tag{62}$$

When there are multiple observations, the intersection of the constraint sets may result in finer range resolution and dynamic range. This is illustrated in Figure 41.

In [42], 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.

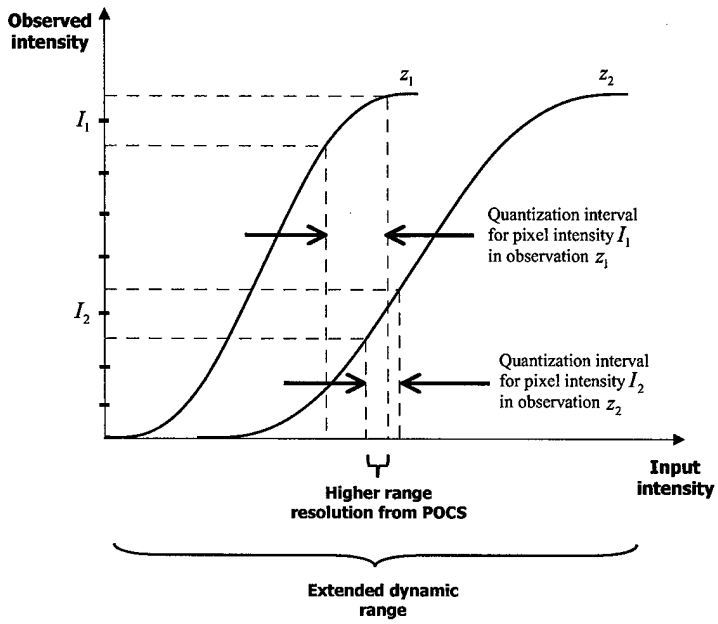


Figure 41: It is possible to increase the gray-scale extent and resolution when there are multiple observations of different range spans.

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 (61) 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 $T_u\left(z_i(m_1, m_2)\right)$, 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 $T_u\left(z_i(m_1, m_2)\right)$. After some simple

algebra, the projection operator onto a constraint set $C\left[z_i(m_1, m_2)\right]$ can be found as

$$P_{C[z_i(m_1,m_2)]}\left[x(n_1,n_2)\right] = x(n_1,n_2) \; + \; \left\{ \begin{array}{c} \left(\frac{T_u(z_i(\mathbf{m}))-\hat{z}_i(\mathbf{m})-\mu_i}{\eta_i}\right)\bar{h}_i(\mathbf{m};\mathbf{n}), \;\; \hat{z}_i(\mathbf{m}) > T_u\left(z_i(\mathbf{m})\right) \\ 0, \qquad T_l\left(z_i(\mathbf{m})\right) < |\hat{z}_i(\mathbf{m})| < T_u\left(z_i(\mathbf{m})\right) \\ \left(\frac{T_l(z_i(\mathbf{m}))-\hat{z}_i(\mathbf{m})-\mu_i}{\eta_i}\right)\bar{h}_i(\mathbf{m};\mathbf{n}), \;\; \hat{z}_i(\mathbf{m}) < T_l\left(z_i(\mathbf{m})\right) \end{array} \right\}, \tag{63}$$

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 Equation (62), and $\bar{h}_i(\mathbf{m};\mathbf{n})$ is the normalized blurring function:

$$\bar{h}_i(\mathbf{m};\mathbf{n}) \equiv \frac{h_i(m_1, m_2; n_1, n_2)}{\sum\limits_{n_1,n_2} |h_i(m_1, m_2; n_1, n_2)|^2}. \tag{64}$$

4.7.3 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 [32] 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 [48].

4.7.4 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 42(a), 42(b), and 42(c). The image in Figure 42(a) was captured with a relatively short exposure time. In Figure 42(a) the low-contrast details in the lighter regions can be seen better, while in Figure 42(c) the tonal fidelity is higher for the darker regions.

We applied the proposed algorithm to these images. The corners detected for the reference images are depicted in Figure 42(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 43. It have 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 44, 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.

4.7.5 Conclusions

In this project, we presented an image fusion algorithm that improves resolution and extent in both the gray-scale and spatial domains. (In the experiments we did not demonstrate increasing spatial extent, which is a relatively straightforward application.) 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.

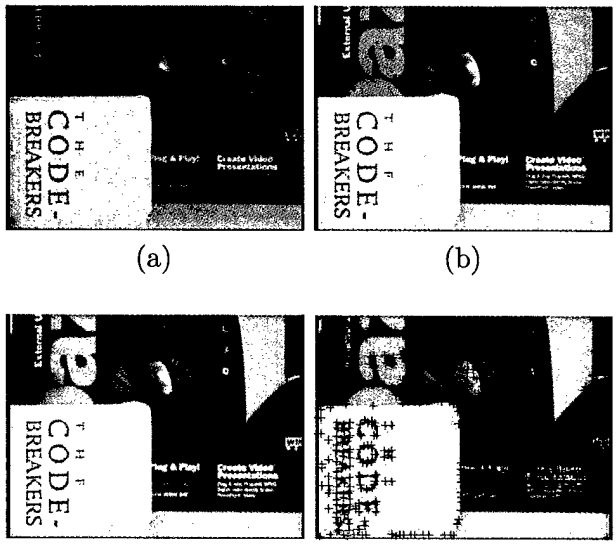


Figure 42: Images from the second sequence. (a) First image. (b) Second image. (c) Third image. (d) Corners detected in the third image.

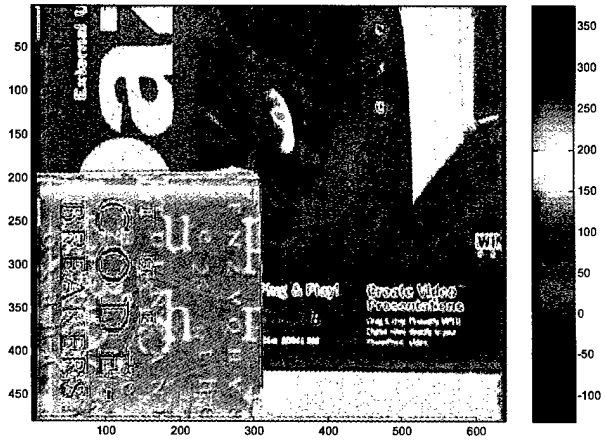


Figure 43: Reconstructed image from the second sequence.

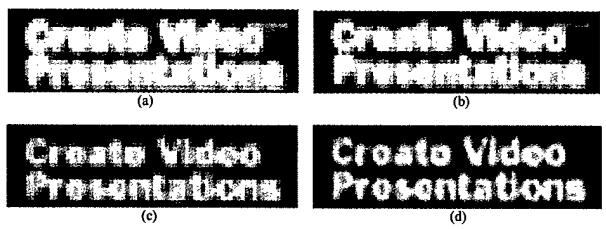


Figure 44: Zoomed regions from the second. (a) First image. (b) Second image. (c) Third image. (d) Reconstructed image scaled to intensity range [0-255].

4.8 Super-Resolution Reconstruction of Hyper-Spectral Images

During the late 1950s, the digital computers began to emerge as a vital tool for dealing with huge amounts of data. Essentially at the same time, significant developments in space technology made the artificial satellites possible. The rapidly growing data processing techniques enabled the use of space imagery for obtaining information and making decisions. After decades of research, space imagery is now a mature field with military and civilian applications, such as target detection, tracking, mineral exploration, and agriculture.

One of the important parameters in a space imagery system is the spatial resolution. There are various effects (atmospheric scattering, secondary illumination changing viewing angle, sensor noise just to name a few) that degrade the acquired image quality. It is obvious that any improvement in spatial resolution will pay off greatly. To improve spatial resolution we make use of the super-resolution techniques together with the information in different wavelengths of the captured spectra, which is made available with the development of the hyper-spectral sensors.

An integral part of our work is to model the hyper-spectral image acquisition process. To get the best results we require our imaging model to be complex enough to incorporate all the effects mentioned above. On the other hand, the model should be kept as simple as possible to avoid computational complexity problems. Currently, we have a candidate model which is being tested under various conditions. The simulation results obtained using this model will also be presented.

4.8.1 Image Acquisition Model

This section starts with some background information about the physical phenomena we are trying to model. In the following section hyper-spectral image acquisition process will be briefly discussed from a physical point of view, together with the atmospheric, environmental and device-dependent effects that influence the overall process. We will describe our hyper-spectral image acquisition model, whereby source image samples are interpreted as aliased and optically blurred linear combinations of the target image's spectral basis. Then a rigorous mathematical formulation of the proposed model will be provided.

- *Background*

To be able explain a hyper-spectral image we require some background information. Any physical object in a scene reflects, absorbs and emits electromagnetic radiation in ways characteristic of its molecular composition and shape. Electro-optical remote sensing is the process of using this fact to obtain information about an object or scene without coming into physical contact with it. If the radiation arriving at the sensor array is measured at a sufficiently high number of wavelengths for every pixel, the resulting spectrum can be used to extract information that cannot be found in images captured by conventional devices (that give no or very little information about the spectral dimension). Spectroscopy is the field concerned with the measurement, analysis and interpretation of such spectra. Another related field, which is actually a branch of spectroscopy, is imaging spectroscopy. It can be explained as combining spectroscopy with methods to acquire spectral information. Hyper-spectral sensors are a class of imaging spectroscopy sensors, for which the waveband of interest is divided into *hundreds* of essentially continuous narrow bands. As the name suggests, the hyper-spectral sensors differ from their predecessors, the multi-spectral sensors; in that the number of bands which can be sensed separately is much higher. (For example AVIRIS Airborne Visible/Infrared Imaging Spectrometer from NASA/JPL has 224 bands.) Hyper-spectral images are the name given to multi-channel images captured by hyper-spectral sensor arrays. They are generally the data type obtained for space imagery application like mining, civil engineering, and military applications such as mine detection and information gathering.

For a given ground pixel, whose dimensions can be in the range of tens of centimeters to tens of meters depending on the spatial resolution of the imaging device, the radiance observed at any particular wavelength is determined by the reflectance of the matter and the solar illumination at that wavelength. However, there exist many important additional effects, which include but are not limited to: imperfect optics of the imaging device, the upwelling solar radiance from atmospheric scattering; the secondary illumination of the object by light reflected from adjacent objects; the scattering and absorption of the reflected radiance by the atmosphere; spatial and spectral aberrations in the sensors, finite sensor dimensions and viewing angle of the sensor array. Characterizing and compensating for these effects is a key step in the exploitation of hyper-spectral imagery. Although the model we propose is very general in the sense that it makes no specific assumptions about the imaging device used and incorporates all effects that have influence on the spatial and spectral resolution of the observed scene, it excludes some physical effects, which are directly related to the sensor characteristics and secondary illumination effects. In the starting phase of this project, we assume that the atmospheric compensation and sensor calibration methods have already been applied and we focus on the image processing side the process.

- *Mathematical Model*

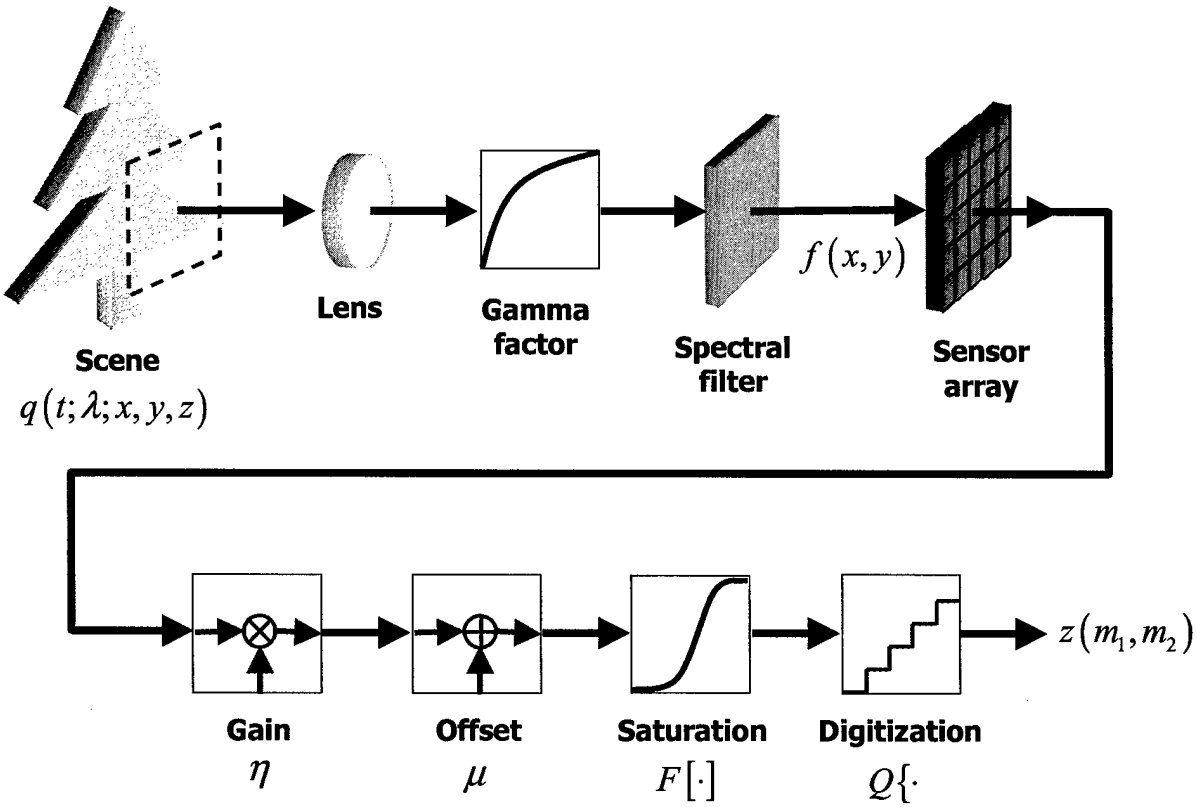


Figure 45: The hyper-spectral image acquisition model.

The block diagram shown in Fig. 45 depicts the system model. The ideal continuous-time, continuous-space and continuous-spectrum video signal, denoted by $f(x_1, x_2, t, \lambda)$, represents the actual input to the imaging device. Ideally, we would like to reconstruct $f(x_1, x_2, t, \lambda)$ from the available observations captured by some hyper-spectral imaging device. However, it is not feasible to reconstruct the continuous signal $f(x_1, x_2, t, \lambda)$. We will overcome this by dimensionality reduction and discretization of the spectral and spatial dimensions.

It is a well-known fact that the spectral reflectance of natural images can be accurately modeled in terms of linear combinations of a relatively small number (generally no more than seven) of reflectance basis functions. These illuminant-independent orthonormal basis functions can be obtained by computing the first P principal components for a large set of natural image reflectances. As a first step in our model, we assume that $f(x_1, x_2, t, \lambda)$ is represented as linear combination of these basis functions. That is, at every pixel location, $f(x_1, x_2, t, \lambda)$ is given by a P dimensional vector where the elements of this vector are the coefficients of corresponding orthonormal basis functions. Once this idea is established, we can proceed with the spatial domain. To deal with the spatial domain, we hypothesize that for each of the P image planes, there exists a corresponding discrete, high-resolution video scene $f_j(n_1, n_2, t)$ $(j = 1, 2, \ldots, P)$ and we seek to reconstruct an image from that signal sampled at a specific time instant, namely $f_j(n_1, n_2, t_r)$. The main assumption here is that the continuous signal $f_j(x_1, x_2, t)$ can be reconstructed from the spatially discrete high-resolution image $f_j(n_1, n_2, t)$ through an ideal reconstruction filter h_r.

In the sections to follow we will model the combination of the image acquisition, spatial and spectral

filtering and sampling. Although the ideas we are trying to convey are rigid, and the path we take to do so is mathematically rigorous, one can easily get confused with the notation used. To leave no room for possible misunderstandings, we end this section with a description of the mathematical notation that is used through out the report. The hyper-spectral image data is best represented in the form of a P-dimensional vector for each pixel where P is the number of spectral bands. Following this convention we write $\underline{f}[\underline{n}] = [f_1[\underline{n}]\ f_2[\underline{n}]\ \ldots\ f_P[\underline{n}]]^T$ to denote the P-dimensional pixel at the location $\underline{n} = [n_1\ n_2]^T$. We use $f_j(x_1, x_2)$ for the spatially continuous target image planes and $f_j[n_1, n_2]$ for the spatially discrete target image planes. Similarly, $g_i(x_1, x_2)$ denotes continuous observation image planes and $g_i(n_1, n_2)$ denotes discrete observation image planes. Hence, any pixel denoted by f—no matter what the subscript or the indices may be—is a target image pixel, and the same applies to g with observation image pixels. Furthermore, at some point it will be necessary to differentiate between high- and low-resolution grid pixels. For this purpose, the high-resolution grid pixels are indexed with $\underline{n} = [n_1\ n_2]^T$ and low-resolution grid pixels are indexed with $\underline{m} = [m_1\ m_2]^T$.

- *Conversion from the Discrete 2D Sequence into a 2D Impulse Train*

The first step in the ideal reconstruction process is the conversion of the discrete signals into 2D impulse trains. Since the following operations are performed on each of the P target image planes, we will drop the subscript j and denote the resulting signal as $f_s(x_1, x_2, t)$:

$$f_s(x_1, x_2, t) = \sum_{n_1=0}^{N} \sum_{n_2=0}^{N} f[n_1, n_2, t]\delta(x_1 - \frac{n_1}{\lambda_1}, x_2 - \frac{n_2}{\lambda_2}). \tag{65}$$

Note that the spatial sampling frequency is normalized for the low-resolution grid so that λ_1 and λ_2 show the increase in the spatial sampling density when we move from the low-resolution image (observation or source) to the high-resolution image (target). In other words, if we assume that the sampling distance in the low-resolution image is 1 unit, then in the same unit, the high-resolution image has λ_1 and λ_2 samples in the horizontal and vertical directions respectively.

- *Ideal Reconstruction Filter*

We implicitly assume that the high-resolution target image (hence its reconstructed version) exists at all times t. Therefore, in the following equations, time index t is suppressed. Keeping this in mind, convolution with the reconstruction filter can be written as

$$f(x_1, x_2) = \iint f_s(x_1 - u_1, x_2 - u_2) h_r(u_1, u_2) du_1 du_2. \tag{66}$$

Substituting $f_s(x_1, x_2)$ from (65) we get

$$f(x_1, x_2) = \iint \sum_{n_1=0}^{N} \sum_{n_2=0}^{N} f[n_1, n_2]\delta(x_1 - u_1 - \frac{n_1}{\lambda_1}, x_2 - u_2 - \frac{n_2}{\lambda_2}) h_r(u_1, u_2)\ du_1 du_2. \tag{67}$$

Assuming convergence we can exchange the order of summations and integrals to write

$$f(x_1, x_2) = \sum_{n_1=0}^{N} \sum_{n_2=0}^{N} f[n_1, n_2] \iint \delta(x_1 - u_1 - \frac{n_1}{\lambda_1}, x_2 - u_2 - \frac{n_2}{\lambda_2}) h_r(u_1, u_2)\ du_1 du_2. \tag{68}$$

Using the sifting property of impulse functional, we get

$$f(x_1, x_2) = \sum_{n_1=0}^{N} \sum_{n_2=0}^{N} f[n_1, n_2] h_r(x_1 - \frac{n_1}{\lambda_1}, x_2 - \frac{n_2}{\lambda_2}), \tag{69}$$

or if we include the suppressed time variable t:

$$f(x_1, x_2, t) = \sum_{n_1=0}^{N} \sum_{n_2=0}^{N} f[n_1, n_2, t] h_r(x_1 - \frac{n_1}{\lambda_1}, x_2 - \frac{n_2}{\lambda_2}). \tag{70}$$

• *Spectral Representation with Predetermined Basis Functions*

We assume that the basis functions are predetermined by applying the principal component analysis (PCA) on training data and selecting the first P principle components. If we denote the continuous signal as $f_c(x_1, x_2, t, \lambda)$, then we have

$$f_c(x_1, x_2, t, \lambda) = \sum_{j=1}^{P} b_j(\lambda) f_j(x_1, x_2, t). \tag{71}$$

Noting that (70) applies to each of the P target image planes, we can write

$$f_c(x_1, x_2, t, \lambda) = \sum_{j=1}^{P} b_j(\lambda) \left[\sum_{n_1=0}^{N} \sum_{n_2=0}^{N} f_j[n_1, n_2, t] h_r(x_{1,t} - \frac{n_1}{\lambda_1}, x_{2,t} - \frac{n_2}{\lambda_2}) \right]. \tag{72}$$

• *Spatial Filtering (Optical Lens & Sensor Integration Blur Model)*

We use $h_\lambda(x_1, x_2)$ to denote the spatially invariant blur filter at every wavelength λ. This models the imperfect imaging optics (e.g. lens blur) and the unavoidable sensor integration blur due to the finite sensor area. The blur operation can be written as the convolution of the target image planes with the blur filter:

$$f_{c,b}(x_1, x_2, t, \lambda) = \iint h_\lambda(x_1 - \nu_1, x_2 - \nu_2) f_c(\nu_1, \nu_2, t, \lambda) d\nu_1 d\nu_2, \tag{73}$$

where subscript c, b means *continuous and blurred.* We use the motion mapping $\mathbf{M}$ for relating the frames occurred at different times. $\mathbf{M} = (M_1, M_2)$ is defined as

$$\begin{aligned} x_{1,t_r} &= M_1(x_1, x_2, t, t_r), \\ x_{2,t_r} &= M_2(x_1, x_2, t, t_r) \ (t_r > t). \end{aligned} \tag{74}$$

For the case at hand, we can write $f_c(x_{1,tr}, x_{2,tr}, t_r, \lambda) = f_c(\nu_1, \nu_2, t, \lambda)$. Then by using the inverse of the mapping mentioned above, we get

$$\begin{aligned} f_{c,b}(x_1, x_2, t, \lambda) = \iint & h_\lambda(x_1 - M_1^{-1}(x_{1,t_r}, x_{2,t_r}, t, t_r), x_2 - M_2^{-1}(x_{1,t_r}, x_{2,t_r}, t, t_r)) |\mathbf{J}| \\ & \times f_c(x_{1,t_r}, x_{2,t_r}, t_r, \lambda) dx_{1,t_r} dx_{2,t_r}, \end{aligned} \tag{75}$$

where $|\mathbf{J}|$ is the Jacobian of the motion mapping. If we make the definition

$$h_{v,\lambda}(x_1, x_2; x_{1,t_r}, x_{2,t_r}; t; t_r) \triangleq |\mathbf{J}| h_\lambda(x_1 - M_1^{-1}(x_{1,t_r}, x_{2,t_r}, t, t_r), x_2 - M_2^{-1}(x_{1,t_r}, x_{2,t_r}, t, t_r)), \quad (76)$$

we can write $f_{c,b}(x_1, x_2, t, \lambda)$ as follows:

$$f_{c,b}(x_1, x_2, t, \lambda) = \iint h_{v,\lambda}(x_1, x_2; x_{1,t_r}, x_{2,t_r}; t; t_r) f_c(x_{1,t_r}, x_{2,t_r}, t_r, \lambda) dx_{1,t_r} dx_{2,t_r}. \quad (77)$$

Substituting the previously derived $f_c(x_{1,t_r}, x_{2,t_r}, t_r, \lambda)$ into this, we get

$$\begin{aligned} & f_{c,b}(x_1, x_2, t, \lambda) \\ & = \iint h_{v,\lambda}(x_1, x_2; x_{1,t_r}, x_{2,t_r}; t; t_r) \\ & \quad \left[\sum_{j=1}^{P} b_j(\lambda) \sum_{n_1=0}^{N} \sum_{n_2=0}^{N} f_j[n_1, n_2, t_r] h_r(x_{1,t_r} - \frac{n_1}{\lambda_1}, x_{2,t_r} - \frac{n_2}{\lambda_2}) \right] dx_{1,t_r} dx_{2,t_r}. \end{aligned} \quad (78)$$

Again assuming convergence, we can exchange the integrals and summations to obtain

$$\begin{aligned} & f_{c,b}(x_1, x_2, t, \lambda) \\ & = \sum_{j=1}^{P} b_j(\lambda) \sum_{n_1=0}^{N} \sum_{n_2=0}^{N} f_j[n_1, n_2, t_r] \\ & \quad \iint h_{v,\lambda}(x_1, x_2; x_{1,t_r}, x_{2,t_r}; t; t_r) h_r(x_{1,t_r} - \frac{n_1}{\lambda_1}, x_{2,t_r} - \frac{n_2}{\lambda_2}) dx_{1,t_r} dx_{2,t_r}. \end{aligned} \quad (79)$$

To get a simpler form, we make the following definition:

$$\begin{aligned} & h_b(x_1, x_2; n_1, n_2; t; t_r) \\ & \quad \triangleq \iint h_{v,\lambda}(x_1, x_2; x_{1,t_r}, x_{2,t_r}; t; t_r) h_r(x_{1,t_r} - \frac{n_1}{\lambda_1}, x_{2,t_r} - \frac{n_2}{\lambda_2}) dx_{1,t_r} dx_{2,t_r}, \end{aligned} \quad (80)$$

which allows us to write

$$f_{c,b}(x_1, x_2, t, \lambda) = \sum_{j=1}^{P} b_j(\lambda) \sum_{n_1=0}^{N} \sum_{n_2=0}^{N} f_j[n_1, n_2, t_r] h_b(x_1, x_2; n_1, n_2; t; t_r). \quad (81)$$

- *Spectral Filtering*

The spectral response functions, $r_1(\lambda)$, $r_1(\lambda)$ ... $r_Q(\lambda)$ model the hyper-spectral sensors' efficiency at different wavelengths as well as the atmospheric and illuminator-based effects on the spectrum.

$$g_i(x_1, x_2, t) = \int_0^\infty f_{c,b}(x_1, x_2, t, \lambda) r_i(\lambda) \, d\lambda. \quad (82)$$

- *Spatial Domain Sampling*

Next, we must discretize the observations for practical implementation. This is done by sampling the g_i's on a low-resolution grid.

$$
\begin{aligned}
&g_i(m_1, m_2, k) \\
&\quad = \int_0^\infty \left[\sum_{j=1}^{P} b_j(\lambda) \sum_{n_1=0}^{N} \sum_{n_2=0}^{N} f_j[n_1, n_2, t_r] h_b(m_1, m_2; n_1, n_2; t_k; t_r) \right] r_i(\lambda) d\lambda \\
&\quad = \sum_{j=1}^{P} \left[\int_0^\infty b_j(\lambda) r_i(\lambda)\, d\lambda \right] \sum_{n_1=0}^{N} \sum_{n_2=0}^{N} f_j[n_1, n_2, t_r] h_b(m_1, m_2; n_1, n_2; t_k; t_r), \qquad (83)
\end{aligned}
$$

where the second equality follows from the assumption that the integrals and summations converge, and hence we can exchange their order. If we call the integral in brackets as $w_{i,j}$, then we can write

$$
g_i(m_1, m_2, k) = \sum_{j=1}^{P} w_{i,j} \sum_{n_1=0}^{N} \sum_{n_2=0}^{N} f_j[n_1, n_2, t_r] h_b(m_1, m_2; n_1, n_2; t_k; t_r), \qquad (84)
$$

or in matrix form, with $Q = 31$ and $P = 6$,

$$
\begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_{31} \end{bmatrix} = \begin{bmatrix} w_{1,1} & \cdots & w_{1,6} \\ w_{2,1} & \cdots & w_{2,6} \\ \vdots & \ddots & \vdots \\ w_{31,1} & \cdots & w_{31,6} \end{bmatrix} \begin{bmatrix} f_1[\underline{n}] \cdot h_b(\underline{m}; \underline{n}; t_k; t_r) \\ f_2[\underline{n}] \cdot h_b(\underline{m}; \underline{n}; t_k; t_r) \\ \vdots \\ f_6[\underline{n}] \cdot h_b(\underline{m}; \underline{n}; t_k; t_r) \end{bmatrix}
$$

where we made the following substitution to get a simpler looking equality:

$$
f_j[\underline{n}] \cdot h_b(\underline{m}; \underline{n}; t_k; t_r) \triangleq \sum_{n_1=0}^{N} \sum_{n_2=0}^{N} f_j[n_1, n_2, t_r] h_b(m_1, m_2; n_1, n_2; t_k; t_r). \qquad (85)
$$

- *Additive Noise*

$v(m_1, m_2, k)$ models the total effect of all possible noise sources that exist through out the whole acquisition process.

4.8.2 Solving the Inverse Problem

Given the model presented in the previous section, the inverse problem can be stated as *finding the target image that is in as much agreement as possible with the observations.* Here *being in agreement* deserves some explanation. When we say the candidate target image is in agreement with the observations we mean that if we apply the linear, time and space-varying (LTSV) filter h_b in (84) to the candidate target image, the resulting synthetic observation image is close to the real observations captured by the imaging device under consideration. There exist many ways to solve this problem, each with its advantages and disadvantages. For example, we can try to minimize the squared error between the observed images and the synthetically produced observation images by using the well-studied least squares methods. The drawback of this approach is that it requires the computation of inverses of large matrices, which is in most cases difficult to compute. A widely

preferred alternative is the iterative set-theoretic methods. In this paper, we will use POCS method to solve the inverse problem addressed above.

Regardless of the method used to solve for the target image, a good share of the total effort is put into calculating the LTSV blur filter h_b. From (80), we can see that h_b has a complex structure. It depends on the reconstruction and spatial blur filter as well as the motion in the scene. Due to these dependencies, the computation of h_b can easily get involved. Furthermore, h_b is only valid where the motion is accurately modelled, making the precision of the motion vectors used in the computations extremely important.

In many cases, computational load or real-time implementation specifications renders a direct calculation impossible and we are required to approximate h_b. A good understanding of the LTSV filtering operation given in (84) is helpful in precisely approximating h_b. For this reason, we will study two special cases, namely single frame and multiple frames with translational motion. It will turn out that, in these cases h_b is fairly easy to compute and has a nice interpretation which sheds light on how we can approximate it.

- *Single Frame Case*

In the single frame case, we have only one hyper-spectral observation, which is a set of monochromatic images of the same scene captured at different wavelengths. Our aim is to reconstruct the high-resolution target image planes $f_j(x_1, x_2, \lambda)$ for $j = 1, 2, \ldots, P$, from which we can obtain the spatially and spectrally continuous target image by using the spectral basis functions. Note that the time dependency is dropped since we are working with only one observation. Following the same steps as in the previous sections one can show that for single observation case, $f_{c,b}(x_1, x_2, \lambda)$ is given by

$$f_{c,b}(x_1, x_2, \lambda) = \sum_{j=1}^{P} b_j(\lambda) \sum_{n_1=0}^{N} \sum_{n_2=0}^{N} f_j[n_1, n_2] \iint h_\lambda(x_1 - \epsilon_1, x_2 - \epsilon_2) h_r(\epsilon_1 - \frac{n_1}{\lambda_1}, \epsilon_2 - \frac{n_2}{\lambda_2})\, d\epsilon_1 d\epsilon_2. \tag{86}$$

If we make the definition

$$\gamma(x_1, x_2) \triangleq h_\lambda(x_1, x_2) * h_r(x_1, x_2), \tag{87}$$

where $*$ denotes convolution, we can see that

$$f_{c,b}(x_1, x_2, \lambda) = \sum_{j=1}^{P} b_j(\lambda) \sum_{n_1=0}^{N} \sum_{n_2=0}^{N} f_j[n_1, n_2] \gamma(x_1 - \frac{n_1}{\lambda_1}, x_2 - \frac{n_2}{\lambda_2}). \tag{88}$$

Continuing with spectral filtering and sampling on a low-resolution grid gives

$$g_i(m_1, m_2) = \sum_{j=1}^{P} w_{i,j} \sum_{n_1=0}^{N} \sum_{n_2=0}^{N} f_j[n_1, n_2] \gamma(m_1 - \frac{n_1}{\lambda_1}, m_2 - \frac{n_2}{\lambda_2}). \tag{89}$$

- *Multi-Frame Translational Motion Case*

In the translational motion case, we have many observations but the motion in the observed scene is constrained to be global translational motion. This type of motion can be incorporated into the

model by letting

$$M_1(x_1, x_2, t, t_r) = x_1 + \delta_1(t_r - t),$$
$$M_2(x_1, x_2, t, t_r) = x_2 + \delta_2(t_r - t). \tag{90}$$

Using these motion mappings and proceeding as in the single frame case, we obtain

$$f_{c,b}(x_1, x_2, t, \lambda) = \sum_{j=1}^{P} b_j(\lambda) \sum_{n_1=0}^{N} \sum_{n_2=0}^{N} f_j[n_1, n_2, t_r]\gamma(x_1 - \frac{n_1}{\lambda_1} + \delta_1(t_r - t), x_2 - \frac{n_2}{\lambda_2} + \delta_2(t_r - t)). \tag{91}$$

Continuing with spectral filtering and sampling on a low-resolution grid gives

$$g_i(m_1, m_2) = \sum_{j=1}^{P} w_{i,j} \sum_{n_1=0}^{N} \sum_{n_2=0}^{N} f_j[n_1, n_2]\gamma(m_1 - \frac{n_1}{\lambda_1} + \delta_1(t_r - t), m_2 - \frac{n_2}{\lambda_2} + \delta_2(t_r - t)). \tag{92}$$

From Eq. (92) we can see that the same interpretation given in the previous section applies to the translational motion case with a slight change. In this case, the effective blur window is moving at the same speed and in same direction as the global translational motion.

4.8.3 Experimental Results

We designed and conducted several experiments to test the proposed method. In these experiments, we used two hyper-spectral data sets. The first set is from the University of Pennsylvania database located at *http://color.psych.upenn.edu/hyperspectral.*. These images are captured under controlled illumination lab environment and have 31 spectral bands (Q=31). The second data set used in the experiments is a 224 band (Q=224) image of an urban scene captured by AVIRIS. (The exact name of the picture is Moffett Field. For detailed information on the data set see [165]. Since the AVIRIS data includes frequencies far beyond the visible range it is meaningless to try to render RGB images for visual evaluation. Therefore, to present visual results we select some specific spectral band (for all the visual results presented in this report the hundredth band is used) and demonstrate the images corresponding to this band. To simulate the hyper-spectral imaging process, these high-quality images are blurred and downsampled in both spatial and spectral domains. The resulting observations have 15 and 112 spectral bands, respectively, and are half the size of the originals spatially. The proposed reconstruction technique was tested under translational and affine motion scenarios. Before giving the detailed simulation results we briefly discuss these motion models and their relevance to the aerial and space imagery. As a notational convenience, we will denote the two dimensional displacement as $\boldsymbol{d}_{\tau,t}(\boldsymbol{x}) = \boldsymbol{x}(\tau) - \boldsymbol{x}(t)$ where $\boldsymbol{x}(t) = (x(t), y(t))^T$ is the 2D position at time t.

• *Translational Motion:* The translational motion model is limited to simple shifts in the image and the displacement vector can be written as follows:

$$\boldsymbol{d}(\boldsymbol{x}) = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} \tag{93}$$

• *Affine Motion:* The six parameter affine motion model can handle shifts, rotations and zooming. The displacement vector depends on the specific pixel location and can be written as:

$$\boldsymbol{d}(\boldsymbol{x}) = \begin{bmatrix} k_1 & k_2 \\ k_4 & b_5 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} k_3 \\ k_6 \end{bmatrix} \tag{94}$$

Considering the conditions under which the hyperspectral image data is acquired, one can be convinced that these are relevant and realistic motion models. For each motion scenario we have three different blur/downsample configurations:

Case 1

- 5×5 Gaussian spatial blur filter with a variance of two.
- Gaussian spectral blur filter with unit variance.
- Downsampling ratio is four in both vertical and horizontal directions.
- For the multi-frame case eight observation images are used.

Case 2

- 5×5 Gaussian spatial blur filter with a variance of two.
- Gaussian spectral blur filter with unit variance.
- Downsampling ratio is two in both vertical and horizontal directions.
- For the multi-frame case four observation images are used.

Case 3

- 3×3 Gaussian spatial blur filter with unit variance
- Gaussian spectral blur filter with unit variance.
- Downsampling ratio is two in both vertical and horizontal directions.
- For the multi-frame case two observation images are used.

The motion vectors are calculated by applying the block matching algorithm [162] on the properly upsampled images. Visual results are presented in Figures 52, 53 and 54 and numerical results are presented in the following section. All the images are RGB rendered from the hyperspectral outputs for qualitative comparison. Close examination of these figures shows that the spatial resolution has been improved during the reconstruction.

We provide the following simulation results to demonstrate the proposed method under the three scenarios mentioned above together with the results of bilinearly interpolating the separate spectral bands. All the results given in the tables below are $PSNR$ values in dB.

From the tables we can see that the proposed method with multiple observations performs much better than bilinear interpolation.

Numerical results in terms of two different fidelity measures are presented below. These measures are $PNSR$ and *band-averaged* $PNSR$. In this paper $PSNR$ is defined as

$$PSNR = 10\log_{10}\left(\frac{S_{peak}}{MSE}\right) dB \tag{95}$$

where S_{peak} stands for the *peak signal power*, and *band-averaged* $PNSR$ is defined as

$$APSNR = 10\log_{10}\left(\frac{\frac{1}{Q}\sum_{i=1}^{Q} S_{peak,i}}{MSE}\right) dB \tag{96}$$

where $S_{peak,i}$ stands for the *peak signal power in the* i^{th} *spectral band.* Since the data we work on is *not* quantized, the maximum signal value is not fixed. The *band-averaged* $PNSR$ ($APSNR$), for which the numerator is calculated as the average of the peak signal powers of all bands, is selected to compensate for this fact.

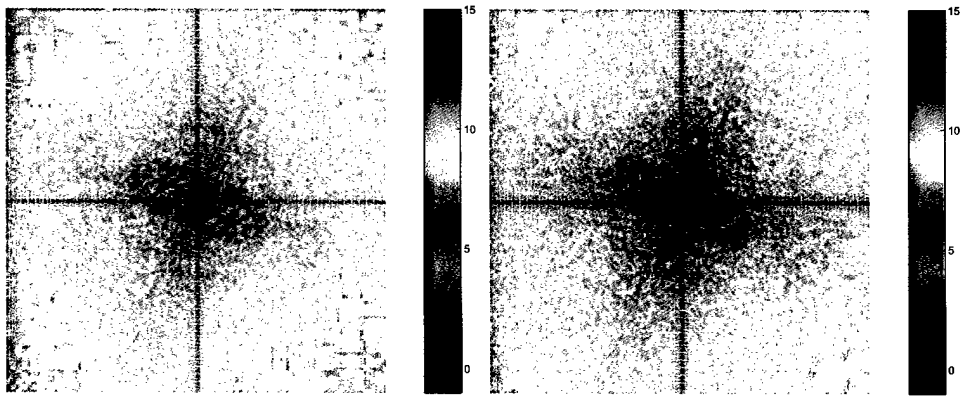
AVIRIS Reflectance Data - 1				
	Bilinear interpolation	Single-cube POCS (no motion)	Multi-cube POCS (translational)	Multi-cube POCS (affine)
— Case 1 —				
PSNR	23.9348	24.3563	25.2773	24.9631
APSNR	21.1726	21.5942	22.5152	22.1395
— Case 2 —				
PSNR	22.9235	23.0498	24.5638	24.2294
APSNR	20.1614	20.2877	21.8017	21.6035
— Case 3 —				
PSNR	22.8220	23.1218	24.1632	23.9102
APSNR	20.0599	20.3597	21.4011	21.0567

AVIRIS Reflectance Data - 2				
	Bilinear interpolation	Single-cube POCS (no motion)	Multi-cube POCS (translational)	Multi-cube POCS (affine)
— Case 1 —				
PSNR	24.2661	24.5539	25.3473	25.1096
APSNR	18.4326	18.7204	19.5138	19.2319
— Case 2 —				
PSNR	23.5391	23.6077	24.7463	24.5771
APSNR	17.6878	17.7564	18.8950	18.6409
— Case 3 —				
PSNR	23.4939	23.6598	24.4973	24.2079
APSNR	17.6426	17.8085	18.5560	18.3508

AVIRIS Radiance Data - 1				
	Bilinear interpolation	Single-cube POCS (no motion)	Multi-cube POCS (translational)	Multi-cube POCS (affine)
— Case 1 —				
PSNR	36.1382	38.3745	42.8398	42.5970
APSNR	28.9405	31.1768	35.6421	35.3455
— Case 2 —				
PSNR	31.9181	32.2407	39.3321	39.0261
APSNR	24.7204	25.0430	32.1344	31.8971
— Case 3 —				
PSNR	31.7053	32.5597	37.1517	36.9012
APSNR	24.5076	25.3620	29.9540	29.6941

AVIRIS Radiance Data - 2				
	Bilinear interpolation	Single-cube POCS (no motion)	Multi-cube POCS (translational)	Multi-cube POCS (affine)
— Case 1 —				
PSNR	35.9959	37.6029	40.5035	40.3064
APSNR	29.1380	30.7451	33.6456	33.3822
— Case 2 —				
PSNR	32.4086	32.7777	37.8406	37.5928
APSNR	25.5500	25.9190	30.9820	30.7559
— Case 3 —				
PSNR	32.1307	33.1518	36.0935	35.8842
APSNR	25.2721	26.2932	29.2349	28.9151

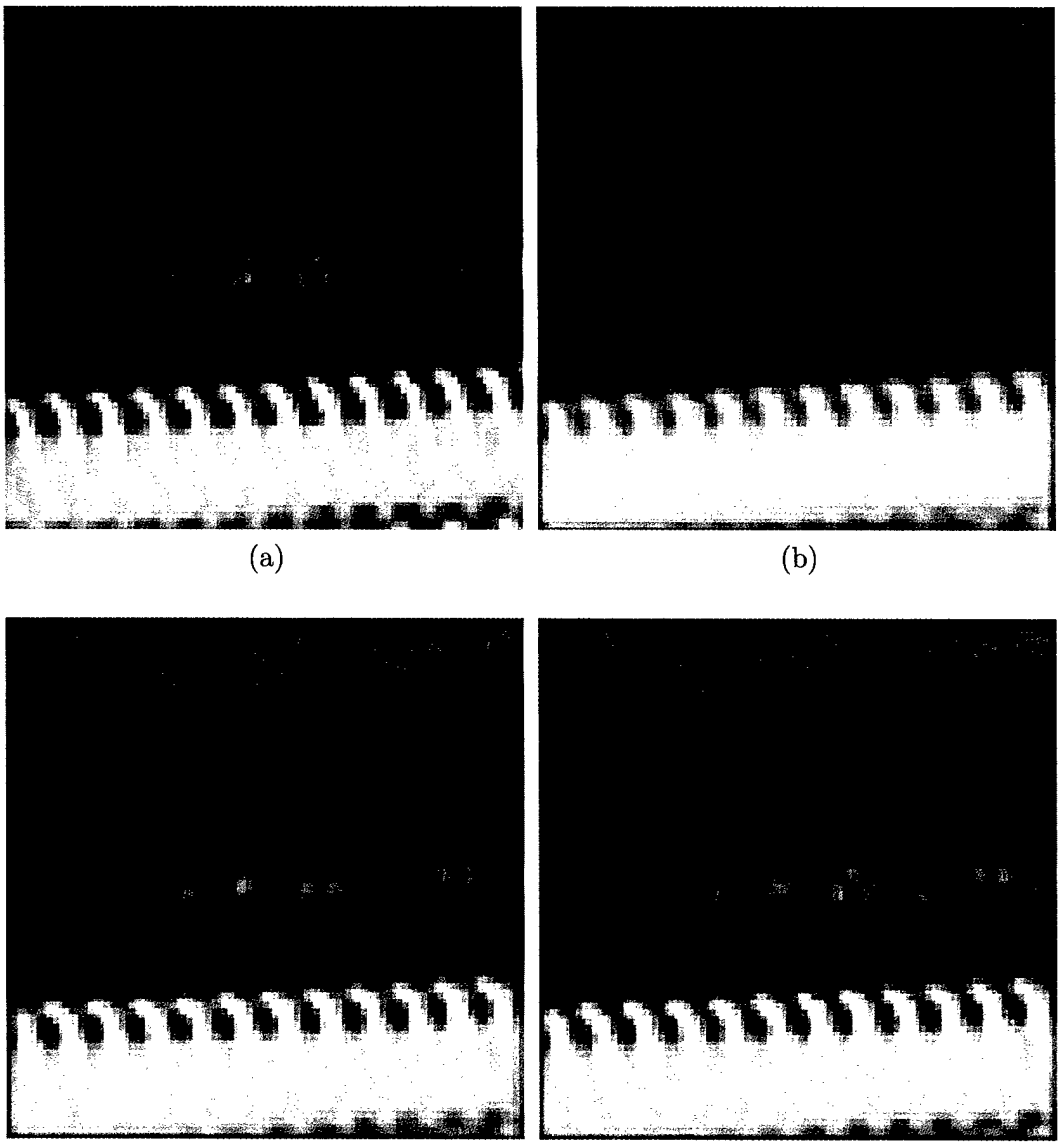
From the tables above we can see that the proposed method even with a single source cube performs better than bilinear interpolation. Using multiple cubes further improves the results, thus pointing out the advantage of fusing the information present across overlapping sources. Visual results presented in Figures (52), (53), (54)and (55) also confirm the improvement seen in $PSNR$ and $APSNR$ values. A careful inspection of the tables also reveals that the $PSNR$ and $APSNR$ values for the Moffett Field reflectance images are 10 to 15 deciBels lower than the values for the radiance images. This offset is caused by the high frequency components present in the reflectance images. Figure (46) shows the 2D FFT of some specific spectral bands of the test images. For the Moffett Field images the hundredth spectral band is used and for the BearFruitGray images it is the tenth spectral band. Note that the values of the color-map bars are in (natural) logarithmic scale. From these plots we can clearly see that the reflectance images have larger high frequency components compared to the radiance images. When filtered with a low-pass filter such as the blur filters we use to obtain our observation images, these components are heavily degraded if not totally lost. This lost information is the main reason for the offset between the radiance and reflectance images.



(a) Radiance 1 (b) Reflectance 1

Figure 46: Frequency contents of some specific bands of the test images. For the Moffett Field images the hundredth spectral band is used.

4.8.4 Figures



(a) (b)

(c) (d)

Figure 47: Experimental results. (a) Original image, (b) One of the observations, (c) Single-frame reconstruction result, (d) Multi-frame reconstruction result. (Four observations used.)

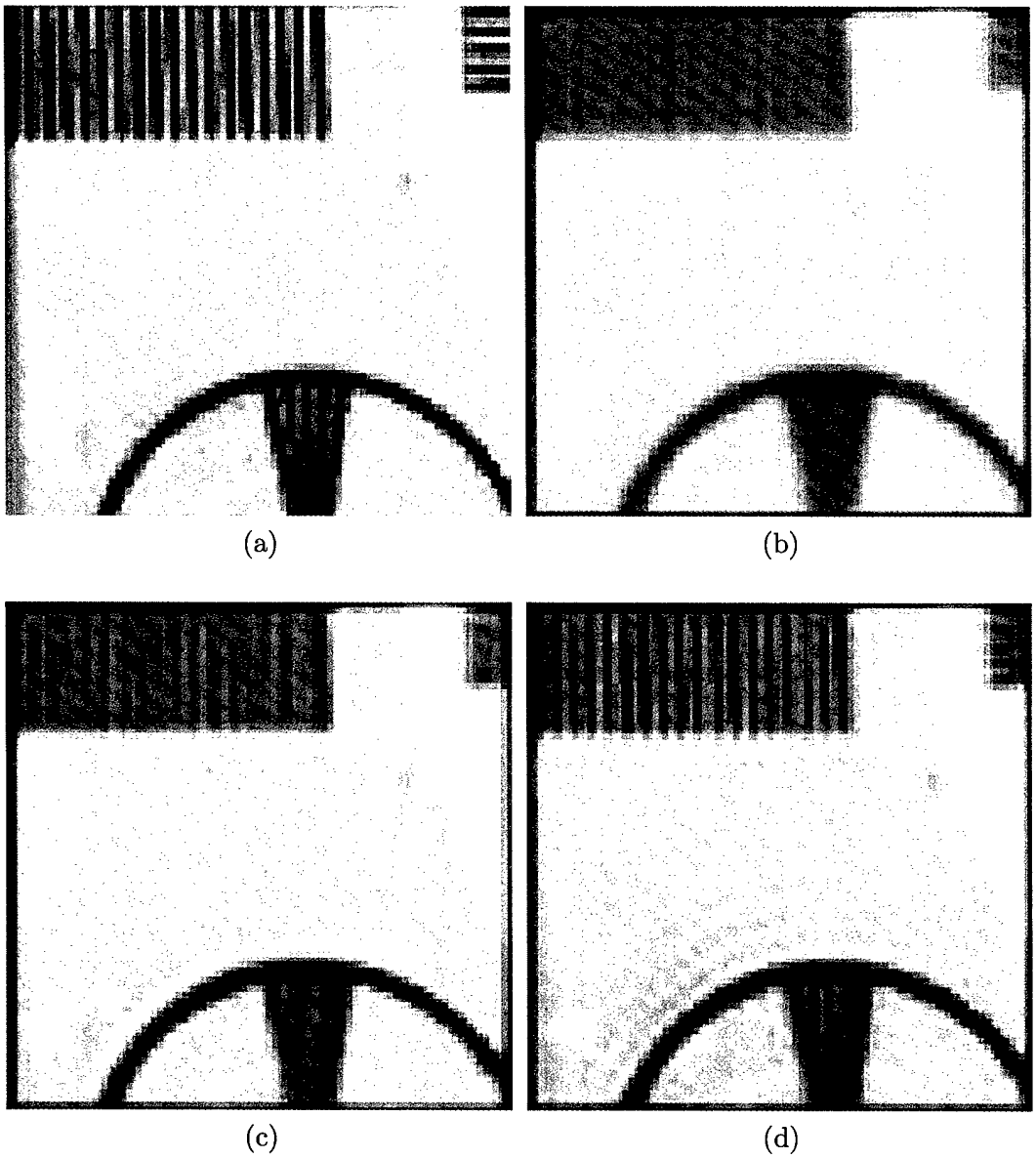
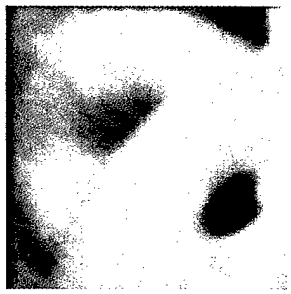


Figure 48: Experimental results. (a) Original image, (b) One of the observations, (c) Single-frame reconstruction result, (d) Multi-frame reconstruction result. (Four observations used.)

(a) Original

(b) Bilinear

(c) Multi frame



(d) Bilinear

(e) Multi frame

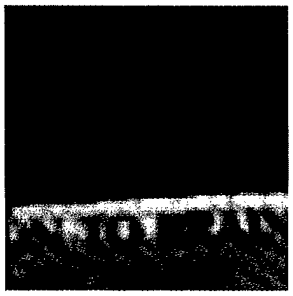
(f) Bilinear

(g) Multi frame

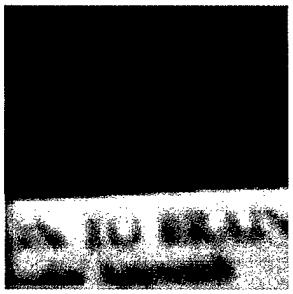
Figure 49: Results for test image chosen from block 31.

(a) Original

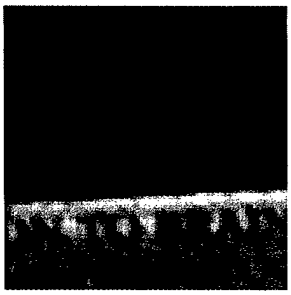
(b) Bilinear



(c) Multi frame



(d) Bilinear

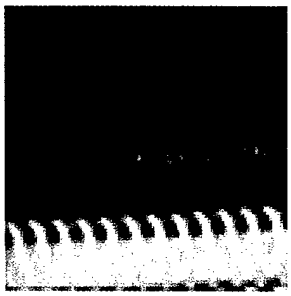


(e) Multi frame

(f) Bilinear

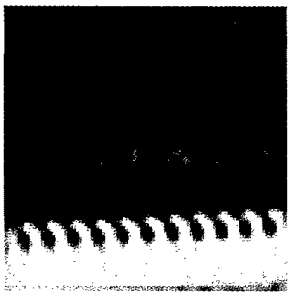
(g) Multi frame

Figure 50: Results for test image chosen from block 34.

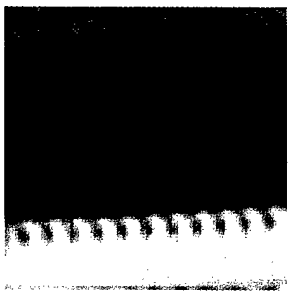


(a) Original

(b) Bilinear

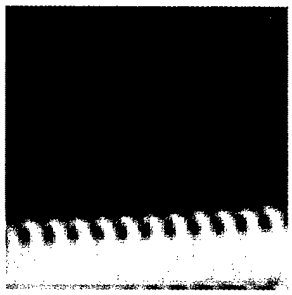


(c) Multi frame

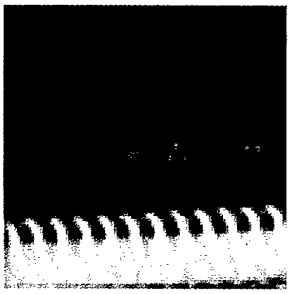


(d) Bilinear

(e) Multi frame



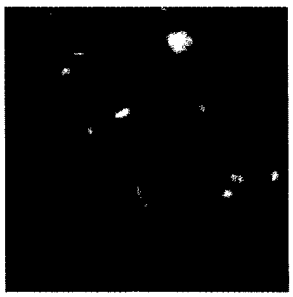
(f) Bilinear



(g) Multi frame

Figure 51: Results for test image chosen from block 44.

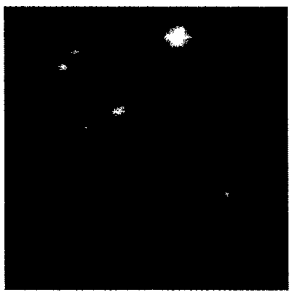
(a) Original



(b) Bilinear

(c) Single cube

(d) Multi cube

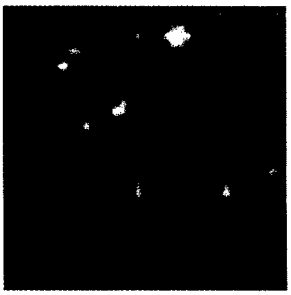


(e) Bilinear

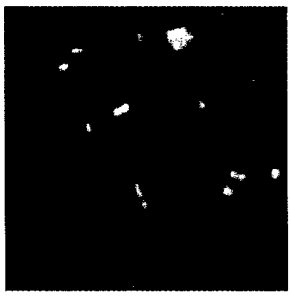
(f) Single cube

(g) Multi cube

(h) Bilinear



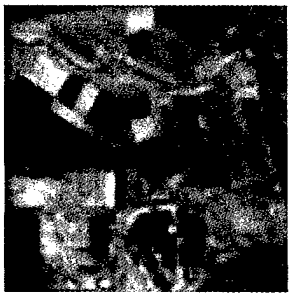
(i) Single cube



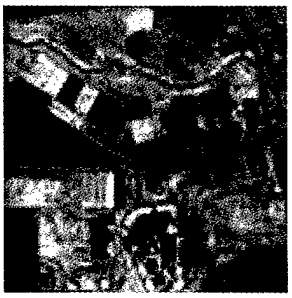
(j) Multi cube

Figure 52: Results for the first reflectance test image extracted from 224-band Moffett Field.

(a) Original

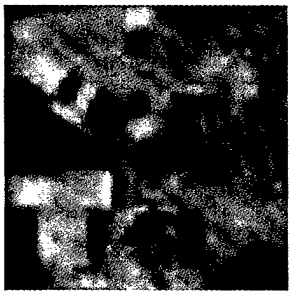


(b) Bilinear

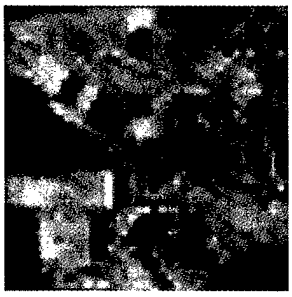


(c) Single cube

(d) Multi cube

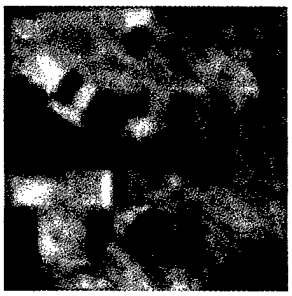


(e) Bilinear

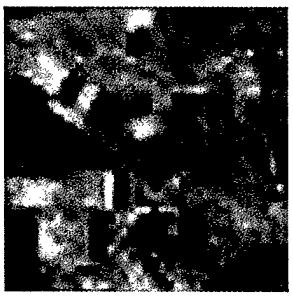


(f) Single cube

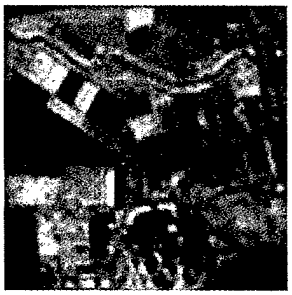
(g) Multi cube



(h) Bilinear



(i) Single cube



(j) Multi cube

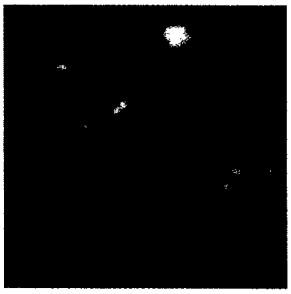
Figure 53: Results for the second reflectance test image extracted from 224-band Moffett Field.

(a) Original

(b) Bilinear

(c) Single cube

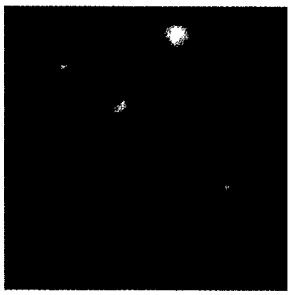
(d) Multi cube



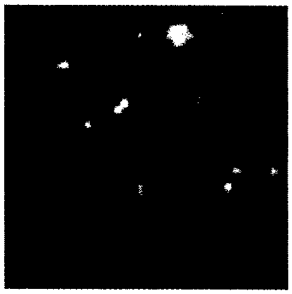
(e) Bilinear

(f) Single cube

(g) Multi cube



(h) Bilinear



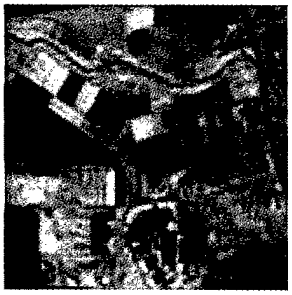
(i) Single cube

(j) Multi cube

Figure 54: Results for the first radiance test image extracted from 224-band Moffett Field.

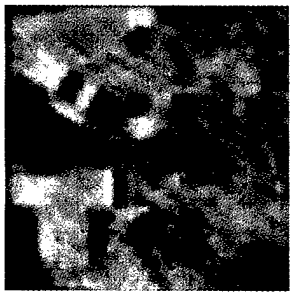
(a) Original

(b) Bilinear

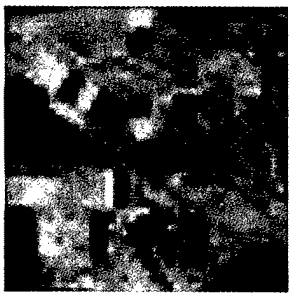


(c) Single cube

(d) Multi cube

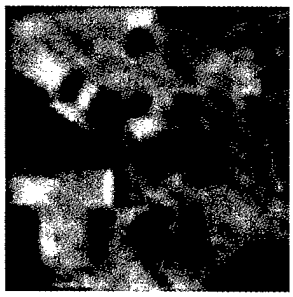


(e) Bilinear

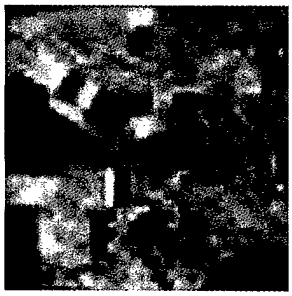


(f) Single cube

(g) Multi cube



(h) Bilinear



(i) Single cube

(j) Multi cube

Figure 55: Results for the second radiance test image extracted from 224-band Moffett Field.

References

[1] H. Gharavi and M. Mills, "Blockmatching motion estimation algorithms–new results," vol. 37, pp. 649–651, 1990.

[2] B. Natarajan, V. Bhaskaran, and K. Konstantinides, "Low-complexity block-based motion estimation via one-bit transforms," vol. 7, pp. 702–706, 1997.

[3] Y. J. Baek, H. S. Oh, and H. K. Lee, "An efficient block-matching criterion for motion estimation and its vlsi implementation," *IEEE Trans. Consumer Electronics*, vol. 42, pp. 885–892, 1996.

[4] S.Lee, J.-M.Kim, and S.-I. Chae, "New motion estimation algorithm using adaptively quantized low bit-resolution image and its vlsi architecture for mpeg2 video encoding," vol. 8, pp. 734–744, 1998.

[5] Y. Altunbasak, A. J. Patti, and O. King, "On the global parametric motion estimation with illumination and geometric lens distortion correction," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, October 1999, pp. 686–690.

[6] S. Baker and T. Kanade, "Hallucinating faces," in *Fourth International Conf. Automatic Face and Gesture Recognition*, March 2000.

[7] M. Berthod, H. Skekarforoush, M. Werman, and J. Zerubia, "Reconstruction of high resolution 3d visual information using subpixel camera displacements," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, 1994, pp. 654–657.

[8] S. Borman and R. L. Stevenson, "Simultaneous multi-frame map super-resolution video enhancement using spatio-temporal priors," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, 1999, pp. 469–473.

[9] N. K. Bose, H. C. Kim, and H. M. Valenzuela, "Recursive total least squares algorithm for image reconstruction from noisy, undersampled frames," *Multidimensional Systems and Signal Processing*, vol. 4, pp. 253–268, 1993.

[10] 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.

[11] ——, "Synthesizing a panoramic scene with a common exposure via the simultaneous registration of images," in *FCRAR 2002*, May 2002.

[12] D. Capel and A. Zisserman, "Automated mosaicing with super-resolution zoom," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, June 1998, pp. 885–891.

[13] ——, "Super-resolution enhancement of text image sequences," in *Proc. IEEE Int. Conf. Pattern Recognition*, vol. 1, 2000, pp. 600–605.

[14] ——, "Super-resolution from multiple views using learnt image models," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, December 2001.

[15] T. E. Boult, M.-C. Chiang, and R. J. Micheals, *Super-Resolution via Image Warping.* Boston: Kluwer Academic Publishers, 2001.

[16] P. Cheeseman, B. Kanefsky, and R. Hanson, "Super-resolved surface reconstruction from multiple images," *Technical Report, NASA*, January 1993.

[17] D. Chen and R. R. Schultz, "Extraction of high-resolution video stills from mpeg image sequences," in *Proc. IEEE Int. Conf. Image Processing*, vol. 2, 1998, pp. 465–469.

[18] T. Sim, S. Baker, and M. Bsat, "The cmu pose, illumination, and expression (pie) database of human faces," The Robotics Institute, Carnegie Mellon University, Tech. Rep. CMU-RI-TR-01-02, 2001.

[19] P. L. Combettes, "The foundations of set theoretic estimation," *Proc. of the IEEE*, vol. 81, no. 2, pp. 182–208, February 1993.

[20] P. E. Devevec, C. J. Taylor, and J. Malik, "Modeling and rendering architecture from photographs," in *SIGGRAPH*, 1996, pp. 11–20.

[21] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy and undersampled measured images," *IEEE Trans. Image Processing*, vol. 6, no. 12, pp. 1646–1658, December 1997.

[22] ——, "Superresolution restoration of an image sequence: adaptive filter approach," *IEEE Trans. Image Processing*, vol. 8, no. 3, pp. 387–395, March 1999.

[23] ——, "Super-resolution reconstruction of image sequences," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 9, pp. 817–834, September 1999.

[24] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Generative models for recognition under variable pose and illumination," in *Proc. IEEE Conf. Face and Gesture Recognition*, 2000, pp. 277–284.

[25] H. Ghassemian, "Multi-sensor image fusion using multirate filter banks," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, 2001, pp. 846–849.

[26] R. B. Gomez, A. Jazaeri, and M. Kafatos, "Wavelet-based hyperspectral an multispectral image fusion," in *2001 SPIE OE/Aerospace Sensing, Geo-Spatial Image and Data Exploitation II*, April 2001.

[27] B. K. Gunturk, Y. Altunbasak, and R. M. Mersereau, "Color plane interpolation using alternating projections," *IEEE Trans. Image Processing*, vol. 11, no. 9, pp. 997–1013, September 2002.

[28] ——, "Super-resolution reconstruction of compressed video using transform-domain statistics," *IEEE Trans. Image Processing*, to appear.

[29] B. K. Gunturk, A. U. Batur, Y. Altunbasak, R. M. Mersereau, and M. H. H. III, "Eigenface-domain super-resolution for face recognition," *IEEE Trans. Image Processing*, to appear.

[30] ——, "Eigenface-based super-resolution for face recognition," in *Proc. IEEE Int. Conf. Image Processing*, to appear in 2002.

[31] P. Hallinan, "A low dimensional representation of human faces for arbitrary lighting conditions," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, 1994, pp. 995–999.

[32] C. G. Harris and M. Stephens, "A combined corner and edge detector," in *Fourth Alvey Vision Conference*, 1988, pp. 147–151.

[33] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint map registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Trans. Image Processing*, vol. 6, no. 12, pp. 1621–1633, December 1997.

[34] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical Models and Image Processing*, vol. 53, pp. 231–239, May 1991.

[35] ——, "Motion analysis for image enhancement: Resolution, occlusion, and transparency," *J. of Visual Communications and Image Representation*, vol. 4, pp. 324–335, December 1993.

[36] T. Jebara, A. Azarbayejani, and A. Pentland, "3d structure from 2d motion," *IEEE Signal Processing Magazine*, vol. 16, no. 3, pp. 66–84, May 1999.

[37] S. P. Kim, N. K. Bose, and H. M. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframes," *IEEE Trans. Acoust. Speech Sign. Proc.*, vol. 38, pp. 1013–1027, June 1990.

[38] S. P. Kim and W. Su, "Recursive high-resolution reconstruction of blurred multiframe images," *IEEE Trans. Image Processing*, vol. 2, pp. 534–539, October 1993.

[39] C. Liu, H.-Y. Shum, and C.-S. Zhang, "A two-step approach to hallucinating faces: Global parametric model and local nonparametric model," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, December 2001.

[40] S. Mann and R. W. Picard, "Virtual bellows: constructing high quality stills from video," in *Proc. IEEE Int. Conf. Image Processing*, November 1994, pp. 13–16.

[41] T. Komatsu, K. Aizawa, and T. Saito, "Very high resolution imaging scheme with multiple different-aperature cameras," *Signal Processing: Image Communications*, vol. 4, pp. 324–335, December 1993.

[42] 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.

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

[44] A. M. Martinez and R. Benavente, "The ar face database," Proc. CVC Technical Report No: 24, Tech. Rep., July 1998.

[45] A. J. Patti and Y. Altunbasak, "Super-resolution image estimation for transform coded video with application to mpeg," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, 1999, pp. 179–183.

[46] ——, "Artifact reduction for set theoretic super resolution image reconstruction with edge adaptive constraints and higher-order interpolants," *IEEE Trans. Image Processing*, vol. 10, no. 1, pp. 179–186, January 2001.

[47] F. Pedersini, A. Sarti, and S. Tubaro, "Multi-camera systems," *IEEE Signal Processing Magazine*, vol. 16, no. 3, pp. 55–65, May 1999.

[48] 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.

[49] M. A. Robertson and R. L. Stevenson, "Dct quantization noise in compressed images," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, 2001, pp. 185–188.

[50] ——, "Dct quantization noise in compressed images," in *SPIE Visual Communications and Image Processing*, vol. 4310, 2001, pp. 21–29.

[51] ——, "Temporal resolution enhancement in compressed video sequences," in *EURASIP Journal on Applied Signal Processing: Special Issue on Nonlinear Signal Processing*, December 2001.

[52] C. A. Segall, R. Molina, A. K. Katsaggelos, and J. Mateos, "Reconstruction of high-resolution image frames from a sequence of low-resolution and compressed observations," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 2, 2002, pp. 1701–1704.

[53] 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.

[54] R. R. Schultz, "Extraction of high-resolution video stills from mpeg image sequences," in *Proc. IEEE Int. Conf. Image Processing*, vol. 2, October 1998, pp. 465–469.

[55] H. Shekarforoush and R. Chellappa, "Data-driven multichannel superresolution with application to video sequences," *J. of the Optical Society of America*, vol. 16, no. 3, pp. 481–492, March 1999.

[56] E. Shechtman, Y. Caspi, and M. Irani, "Increasing space-time resolution in video," in *European Conference on Computer Vision (ECCV)*, May 2002.

[57] L. Sirovich and M. Kirby, "Low dimensional procedure for the characterization of human faces," *J. of the Optical Society of America*, vol. 4, no. 3, pp. 519–524, 1987.

[58] C. Srinivas and M. D. Srinath, "A stochastic model-based approach for simultaneous restoration of multiple misregistered images," *SPIE*, vol. 1360, pp. 1416–1427, 1990.

[59] H. Stark and J. W. Woods, *Probability, Random Processes, and Estimation Theory for Engineers.* Englewood Cliffs: Prentice-Hall Inc., 1986.

[60] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays, using convex projections," *J. of the Optical Society of America*, vol. 6, no. 11, pp. 1715–1726, 1989.

[61] R. Szeliski, "Video mosaics for virtual environments," in *Computer Graphics Applications*, March 1996, pp. 22–30.

[62] A. M. Tekalp, M. K. Ozkan, and M. I. Sezan, "High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, March 1992, pp. 169–172.

[63] A. M. Tekalp, *Digital Video Processing.* Prentice Hall, 1995.

[64] B. C. Tom, A. K. Katsaggelos, and N. P. Galatsanos, "Reconstruction of a high resolution image from registration and restoration of low resolution images," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, November 1994, pp. 553–557.

[65] B. C. Tom, K. T. Lay, and A. K. Katsaggelos, "Multi-channel image identification and restoration using the expectation-maximization algorithm," *Optical Engineering, Special Issue on Visual Communications and Image Processing*, vol. 35, no. 1, pp. 241–254, January 1996.

[66] B. C. Tom and A. K. Katsaggelos, "Resolution enhancement of monochrome and color video using motion compensation," *IEEE Trans. Image Processing*, vol. 10, no. 2, pp. 278–287, February 2001.

[67] R. Y. Tsai and T. S. Huang, *Multiframe Image Restoration and Registration, In: Advances in Computer Vision and Image Processing, ed. T.S.Huang. Greenwich, CT.* JAI Press, 1984.

[68] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, 2001, pp. 586–591.

[69] H. Ur and D. Gross, "Improved resolution from subpixel shifted pictures," *CVGIP: Graphical Models and Image Processing*, vol. 54, pp. 181–186, March 1992.

[70] J. E. Adams, "Interactions between color plane interpolation and other image processing functions in electronic photography," *Proc. SPIE Cameras and Systems for Electronic Photography and Scientific Imaging*, vol. 2416, pp. 144–151, February 1995.

[71] J. E. Adams and J. F. Hamilton, "Adaptive color plane interpolation in single color electronic camera," *U.S. Patent 5,506,619*, April 1996.

[72] ——, "Design of practical color filter array interpolation algorithms for digital cameras," *Proc. SPIE Real Time Imaging II*, vol. 3028, pp. 117–125, February 1997.

[73] J. E. Adams, "Design of color filter array interpolation algorithms for digital cameras, part 2," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, 1998, pp. 488–492.

[74] D. R. Cok, "Color imaging array," *U.S. Patent 3,971,065*, July 1976.

[75] E. Chang, S. Cheung, and D. Y. Pan, "Color filter array recovery using a threshold-based variable number of gradients," vol. 3650, pp. 36–43, 1999.

[76] D. R. Cok, "Signal processing method and apparatus for sampled image signals," *U.S. Patent 4,630,307*, February 1984.

[77] ——, "Signal processing method and apparatus for producing interpolated chrominance values in a sampled color image signal," *U.S. Patent 4,642,678*, February 1987.

[78] J. W. Glotzbach, R. W. Schafer, and K. Illgner, "A method of color filter array interpolation with alias cancellation properties," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, 2001, pp. 141–144.

[79] J. F. Hamilton and J. E. Adams, "Adaptive color plane interpolation in single sensor color electronic camera," *U.S. Patent 5,629,734*, May 1997.

[80] R. H. Hibbard, "Apparatus and method for adaptively interpolating a full color image utilizing luminance gradients," *U.S. Patent 5,382,976*, January 1995.

[81] R. Kimmel, "Demosaicing: image reconstruction from ccd samples," *IEEE Trans. Image Processing*, vol. 8, pp. 1221–1228, 1999.

[82] C. A. Laroche and M. A. Prescott, "Apparatus and method for adaptively interpolating a full color image utilizing chrominance gradients," *U.S. Patent 5,373,322*, December 1994.

[83] J. Mukherjee, R. Parthasarathi, and S. Goyal, "Markov random field processing for color demosaicing," *Pattern Recognition Letters*, vol. 22, pp. 339–351, 2001.

[84] D. Taubman, "Generalized wiener reconstruction of images from colour sensor data using a scale invariant prior," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, 2000, pp. 801–804.

[85] H. J. Trussell and R. E. Hartwig, "Mathematics for demosaicking," *IEEE Trans. Image Processing*, vol. 3, no. 11, pp. 485–492, April 2002.

[86] J. A. Weldy, "Optimized design for a single-sensor color electronic camera system," vol. 1071, pp. 300–307, 1988.

[87] X. Wu, W. K. Choi, and P. Bao, "Color restoration from digital camera data by pattern matching," vol. 3018, pp. 12–17, 1997.

[88] D. Bailey, "Super-resolution of bar codes," *Proceedings of the SPIE - The International Society for Optical Engineering*, vol. 3521, pp. 204–213, Nov 1998.

[89] B. Hunt, "Super-resolution of imagery: understanding the basis for recovery of spatial frequencies beyond the diffraction limit," *Proceeding of Information Decision and Control Adelaide, SA, Australia*, pp. 243–248, Feb 1999.

[90] C. Hein, "Real-time implementation of super-resolution imaging algorithm," *Proc. SPIE, Advanced Signal Processing Algorithms, Architectures, and Implementations VIII San Diego, CA, USA*, vol. 3461, pp. 501–511, July 1998.

[91] P.Min-Cheng and A.H.Lettington, "Efficient method for improving poisson map super-resolution," *Electronics Letters*, vol. 35, no. 10, pp. 803–805, May 1999.

[92] A.Lorette, H.Shekarforoush, and J.Zerubia, "Super-resolution with adaptive regularization," *Proceedings of International Conference on Image Processing Santa Barbara, CA, USA*, vol. 1, pp. 169–172, Oct 1997.

[93] R.R.Schultz and R.L.Stevenson, "Bayesian estimation of subpixel-resolution motion fields and high-resolution video stills," *Proceedings of International Conference on Image Processing Santa Barbara, CA, USA*, vol. 3, pp. 62–65, Oct 1997.

[94] L. Xiao-Zheng, M. Kudo, J. Toyama, and M. Shimbo, "Knowledge-based enhancement of low spatial resolution images," *IEICE Transactions on Information and Systems*, vol. E81-D, no. 5, pp. 457–463, May 1998.

[95] I.N.Aizenberg, N.N.Aizenberg, J.Vandewalle, and V.Tavsanoglu, "Solution of the super-resolution problem by multi-valued nonlinear filtering, and its implementation using cellular neural networks," *Fifth IEEE International Workshop on Cellular Neural Networks and their Applications. Proceedings London, UK*, pp. 353–358, Apr 1998.

[96] D.G.Sheppard, K.Panchapakesan, A.Bilgin, B.R.Hunt, M.W.Marcellin, M.P.Fargues, and R.D.Hippenstiel, "Lapped nonlinear interpolative vector quantization and image super-resolution," *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 224–228, Nov 1997.

[97] D.G.Sheppard, B.R.Hunt, and M.W.Marcellin, "Iterative multiframe super-resolution algorithms for atmospheric turbulence-degraded imagery," *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing Seattle, WA, USA*, vol. 5, pp. 2857–2860, May 1998.

[98] C.L.Matson and D.W.Tyler, "Positivity and support: a comparison [image processing]," *Image Processing, Signal Processing, and Synthetic Aperture Radar for Remote London, UK*, vol. 3217, pp. 200–209, Sept 1997.

[99] A.Zomet and S.Peleg, "Applying super-resolution to panoramic mosaics," *Proceedings of WACV98 - Computer Vision Princeton, NJ, USA*, pp. 286–287, Oct 1998.

[100] R.R.Schultz and R.L.Stevenson, "Estimation of subpixel-resolution motion fields from segmented image sequences," *Proc. SPIE, Sensor Fusion: Architectures, Algorithms, and Applications II Orlando, FL, USA*, vol. 3376, pp. 90–101, Apr 1998.

[101] H.Shekarforoush, R.Chellappa, H.Niemann, H.Seidel, and B.Girod, "Multi-channel super-resolution for image sequences with application to airborne video data," *Proceedings of IMDSP'98 10th Image and Multi-dimensional Digital Signal Processing Workshop Alpbach, Austria*, pp. 207–210, July 1998.

[102] M.C.Chiang and T.E.Boult, "Efficient image warping and super-resolution," *Proceedings Third IEEE Workshop on Applications of Computer Vision. WACV'96 Sarasota, FL, USA*, pp. 56–61, Dec 1996.

[103] M.Elad and A.Feuer, "Super-resolution reconstruction of an image," *Proceedings of 19th Convention of Electrical and Electronics Engineers in Israel Jerusalem, Israel*, pp. 391–394, Nov 1996.

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

[105] B. R. Hunt, "Super-resolution of images: algorithms, principles, performance," *International Journal of Imaging Systems and Technology*, vol. 6, no. 4, pp. 297–304, 1995.

[106] M.Elad and A. Feuer, "Super-resolution restoration of continuous image sequence using the lms algorithm," *Eighteenth Convention of Electrical and Electronics Engineers in Israel Tel Aviv, Israel*, vol. p.2.2.5, pp. 1–5, Mar 1995.

[107] S. Chen and D. Donoho, "Examples of basis pursuit," *Proc. SPIE, Wavelet Applications in Signal and Image Processing III San Diego, CA, USA*, vol. 2569, no. pt.2, pp. 564–574, 1995.

[108] H.Shekarforoush, M.Berthod, and J.Zerubia, "3d super-resolution using generalized sampling expansion," *Proceedings International Conference on Image Processing Washington, DC, USA*, vol. 2, pp. 300–303, Oct 1995.

[109] D.Keightley and B.R.Hunt, "A rigid pocs extension to a poisson super-resolution algorithm," *Proceedings International Conference on Image Processing Washington, DC, USA*, vol. 2, pp. 508–511, Oct 1995.

[110] B.Bascle, A.Blake, and A.Zisserman, "Motion deblurring and super-resolution from an image sequence," *Proceedings of Fourth European Conference on Computer Vision. ECCV '96 Cambridge, UK*, vol. 2, pp. 573–582, Apr 1996.

[111] A. Tanaka, H. Imai, M. Miyakoshi, and T. Da-Te, "Enlargement of digital images with multiresolution analysis," *Transactions of the Institute of Electronics, Information and Communication Engineers D-II*, vol. J79D-II, no. 5, pp. 819–825, May 1996.

[112] P.Yuen and B.R.Hunt, "Constraining information for improvements to map image super-resolution," *Proc. SPIE, Image Reconstruction and Restoration San Diego, CA, USA*, vol. 2302, pp. 144–155, July 1994.

[113] E.Fussfeld and Y.Y.Zeevi, "Super-resolution estimation of edge images," *Proc. SPIE, Visual Communications and Image Processing '94 Chicago, IL, USA*, vol. 2308, no. pt.2, pp. 787–798, Sept 1994.

[114] M.S.Nadar, B.R.Hunt, and P.J.Sementilli, "Multigrid techniques and wavelet representations in image super-resolution," *Proc. SPIE, Visual Communications and Image Processing '94 Chicago, IL, USA*, vol. 2308, no. pt.2, pp. 1244–1255, Sept 1994.

[115] E.Fussfeld and Y.Y.Zeevi, "Super-resolution estimation of edge images," *Proceedings of 12th International Conference on Pattern Recognition Jerusalem, Israel*, vol. 1, pp. 11–16, Oct 1994.

[116] H.Fahimi, "A fast new approach to super-resolution," *Applications of Digital Image Processing XVIII San Diego, CA, USA*, vol. 2564, pp. 117–125, July 1995.

[117] E.A.Salerno, J.Vandewalle, R.Boite, M.Moonen, and A.Oosterlinck, "Achieving super-resolution with band-pass images," *Signal Processing VI - Theories and Applications. Proceedings of EUSIPCO-92, Sixth European Signal Processing Conference*, vol. 3, pp. 1417–1420, Aug 1992.

[118] P.J.Sementilli, M.S.Nadar, and B.R.Hunt, "Poisson map super-resolution estimator with smoothness constraint," *Proc. SPIE, Neural and Stochastic Methods in Image and Signal Processing II San Diego, CA*, vol. 2032, pp. 2–13, July 1993.

[119] B. R. Hunt, "Imagery super-resolution: emerging prospects," *Proc. SPIE, Applications of Digital Image Processing XIV San Diego, CA*, vol. 1567, pp. 600–608, July 1991.

[120] S.P.Luttrell, "A bayesian derivation of an iterative autofocus/super-resolution algorithm," *Inverse Problems*, vol. 6, no. 6, 975-996 1990.

[121] J.B.Abbiss, B.J.Brames, and M.A.Fiddy, "On the application of neural networks to the solution of image restoration problems," *Proc. SPIE, High Speed Computing II Los Angeles, CA*, vol. 1058, pp. 138–146, Jan 1989.

[122] S.P.Luttrell, "Bayesian autofocus/super-resolution theory," *IEE Colloquium on 'Role of Image Processing in Defence and Military Electronics*, vol. 1, pp. 1–6, Apr 1990.

[123] L.M.Kani and J.C.Dainty, "Super-resolution using the gerchberg algorithm," *Optics Communications*, vol. 68, no. 1, pp. 11–17, Sept 1988.

[124] I.Overington and P.Greenway, "Theoretical and practical super-resolution performance in simple processing of local form, motion and stereo disparity," *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 728, pp. 10–18, Oct. 1986.

[125] B.C.Tom, A.K.Katsaggelos, G.Ramponi, G.L.Sicuranza, S.Carrato, and S.Marsi, "Resolution enhancement of color video," *Signal Processing VIII, Theories and Applications. Proceedings of EUSIPCO-96, Eighth European Signal Processing Conference*, vol. 1, pp. 145–148, Sept. 1996.

[126] G.Lilienfield and J.W.Woods, "Scalable high definition video coding," *Proceedings of the SPIE - The International Society for Optical Engineering*, vol. 3309, pp. 158–169, Jan. 1997.

[127] R.R.Schultz and R.L.Stevenson, "Estimation of subpixel-resolution motion fields from segmented image sequences," *Proceedings of the SPIE - The International Society for Optical Engineering*, vol. 3376, pp. 90–101, Apr. 1998.

[128] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Robust methods for high-quality stills from interlaced video in the presence of dominant motion," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 2, pp. 328–342, Apr. 1997.

[129] P.E.Eren, M.I.Sezan, and A.M.Tekalp, "Robust region-based high-resolution image reconstruction from low-resolution video," *Proceedings. International Conference on Image Processing*, vol. 1, pp. 709–712, Sept. 1996.

[130] P. E. Eren, M. I. Sezan, and A. M. Tekalp, "Robust, object-based high-resolution image reconstruction from low-resolution," *IEEE Transactions on Image Processing*, vol. 6, no. 10, pp. 1446–1451, Oct. 1997.

[131] L. Ping and Y. mei Wen, "Image resolution improvement by swing ccd array imager in two dimensions and its signal processing," *Proceedings of the SPIE - The International Society for Optical Engineering*, vol. 2308,pt2, pp. 915–920, sept. 1994.

[132] R.R.Schultz and R.L.Stevenson, "Video resolution enhancement," *Proceedings of the SPIE - The International Society for Optical Engineering*, vol. 2421, pp. 23–34, Feb. 1995.

[133] R. R. Schultz and R. L. Stevenson, "Improved definition video frame enhancement," *1995 International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*, vol. 4, pp. 2169–2172, May 1995.

[134] E.R.Doering and J.P.Basart, "Improving the spatial resolution of real-time radiography images using the maximum entropy method," *Conference Record of the Twenty-Fifth Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 1072–1076, Nov. 1991.

[135] S.-J. Yeh and H. Stark, "Iterative and one-step reconstruction from nonuniform samples by convex projections," *J. of the Optical Society of America*, vol. 7, no. 3, pp. 491–499, 1990.

[136] M. I. Sezan and H. J. Trussell, "Prototype image constraints for set-theoretic image restoration," *IEEE Transactions on Signal Processing*, vol. 39, no. 10, pp. 2275–85, 1991.

[137] H. J. Trussell and B. R. Hunt, "Feasible solution in signal restoration," *IEEE Trans. Acoust. Speech Sign. Proc.*, vol. ASSP-32, pp. 201–212, 1984.

[138] R. L. Stevenson, B. E. Schmitz, and E. J. Delp, "Discontinuity preserving regularization of inverse visual problems," *IEEE Trans. Syst., Man, Cybern. no. 3*, vol. 24, pp. 455–469, 1994.

[139] K. Jensen and D. Anastassiou, "Spatial resolution enhancement of images using non-linear interpolation," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Albuquerque, NM, April 3–6, 1990, pp. 2045–2048.

[140] P. L. Combettes, "Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections," *IEEE Trans. Image Processing*, vol. 6, no. 4, pp. 493–506, April 1997.

[141] M. Bierling, "Displacement estimation by hierarchical block matching," in *Proc. SPIE Visual Communications and Image Processing'88*, 1998, pp. 942–951.

[142] H. J. Trussell and M. R. Civanlar, "Feasible solution in signal restoration," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 32, pp. 201–212, 1984.

[143] B. Tom and A. Katsaggelos, "Reconstruction of a high resolution image by simultaneous registration, restoration and interpolation of low resolution images," in *Proc. Int. Conf. Image Processing*, vol. 2, 1995, pp. 539–542.

[144] Y. Altunbasak, A. Patti, and R. Mersereau, "Super-resolution still and video reconstruction from mpeg-coded video," vol. 12, no. 4, pp. 217–226, 2002.

[145] A. Patti, M. Sezan, and A. M. Tekalp, "Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time," vol. 6, no. 8, pp. 1064–1076, 1997.

[146] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," vol. 6, no. 12, pp. 1646–1658, 1997.

[147] P. Combettes, "Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections," vol. 6, no. 4, pp. 493–506, 1997.

[148] L. Maloney, "Evaluation of linear models of surface spectral reflectance with small numbers of parameters," *Jourmal of the Optical Society of America A*, vol. 3, pp. 1673–1683, 1986.

[149] P. E. Eren, M. I. Sezan, and A. M. Tekalp, "Robust, object-based high-resolution image reconstruction from low-resolution," vol. 6, pp. 1446–1451, 1997.

[150] R. L. Stevenson, B. E. Schmitz, and E. J. Delp, "Discontinuity preserving regularization of inverse visual problems," vol. 24, pp. 455–469, 1994.

[151] G. Shaw and D. Manolakis, "Signal processing for hyperspectral image exploitation," vol. 19, pp. 12–16, Jan. 2002.

[152] D. Landgrebe, "Hyperspectral image data analysis," vol. 19, pp. 17–28, Jan. 2002.

[153] P. Combettes, "Convex set theoretic image recovery with inexact projection algorithms," in *2001 IEEE International Conference on Image Processing (ICIP'01) Proceedings*, vol. 1, 2001, pp. 257–260.

[154] Y. Altunbasak and A. Patti, "A maximum a posteriori estimator for high resolution video reconstruction from mpeg video," in *2000 IEEE International Conference on Image Processing (ICIP'00) Proceedings*, vol. 2, 2000, pp. 649–652.

[155] A. Zomet and S. Peleg, "Applying super-resolution to panoramic mosaics," in *Proceedings of the Fourth IEEE Workshop on Applications of Computer Vision (WACV '98)*, 1998, pp. 286–287.

[156] ——, "Efficient super-resolution and applications to mosaics," in *Proceedings of the 15th International Conference on Pattern Recognition (ICPR 2000)*, vol. 1, 2000, pp. 579–583.

[157] ——, "Robust super-resolution," in *Proceedings of the 2001 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2001)*, vol. 1, 2001, pp. 645–650.

[158] M. Irani and S. Peleg, "Super resolution from image sequences," in *Proceedings of the 10th International Conference on Pattern Recognition (ICPR 1990)*, vol. 2, 1990, pp. 115–120.

[159] B. R. Hunt, "Imagery super-resolution: Emerging prospects," in *Proc. SPIE-Applications of Digital Image Processing XIV*, vol. 1567, San Diego, CA, July, 1991, pp. 600–608.

[160] B. Tom and A. Katsaggelos, "Reconstruction of a high resolution image by simultaneous registration, restoration and interpolation of low resolution images," in *Proc. Int. Conf. Image Processing*, vol. 2, 1995, pp. 539–542.

[161] R. R. Schultz and R. L. Stevenson, "Improved definition video frame enhancement," in *Proc. 1995 Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 4, May, 1995, pp. 2169–2172.

[162] A. M. Tekalp, *Digital Video Processing.* Upper Saddle River, NJ 07458: Prentice Hall, 1995.

[163] T. Huang and R. Tsai, *Multiframe Image Restoration and Registration, In: Advances in Computer Vision and Image Processing*, T. S. Huang, Ed. Greenwich, CT: JAI Press Inc, 1984, vol. 1.

[164] D. H. Brainard. Hyperspectral image data. [Online]. Available: http://color.psych.ucsb.edu/ /hyperspectral/

[165] Aviris free data. Jet Propulsion Laboratory, California Institute of Technology. [Online]. Available: http://popo.jpl.nasa.gov/html/aviris.fredata.html