Performance Evaluation of Super-Resolution Reconstruction from Video

Gulcin Caner, A. Murat Tekalp^a and Wendi Heinzelman

Department of Electrical and Computer Engineering, University of Rochester, Rochester,
NY 14627

^a(also with) College of Engineering, Koc University, Istanbul, Turkey
{caner,wheinzel,tekalp}@ece.rochester.edu
www.ece.rochester.edu/people

ABSTRACT

Several algorithms have been proposed to enhance the resolution of a reference image from multiple still images or video that may be captured/stored in raw or compressed format. This paper provides a thorough study of how the performance of the projections onto convex sets (POCS) method is affected by camera parameters, quantization of the pixel-values, motion estimation errors, and quantization in the DCT-domain (when compressed data is used). Experimental results are provided to evaluate the practical applicability of super-resolution reconstruction in various scenarios. It has been observed that the quality of the resolution enhancement depends on the quantization of pixel intensity values in the RGB (uncompressed) or YUV (compressed) domains by the video capture device, as well as the accuracy of the estimated motion parameters between successive frames.

Keywords: Super-resolution, POCS, performance evaluation

1. INTRODUCTION

The spatial resolution of images and video is limited by the imaging device. It is often of interest to obtain a reference image or region of interest at a resolution higher than that of the imaging device. The process of combining information from multiple low-resolution still images with sub-pixel displacements to obtain a higher resolution reference image is called super-resolution (SR) reconstruction in the literature. SR reconstruction has been found very useful in many practical applications, such as medical imaging, forensic imaging, and surveillance systems. Another application area has been printing still images from video.

SR reconstruction techniques can be classified into two main groups, spatial domain methods that work with uncompressed video, and DCT-domain methods that account for the effect of quantization in the DCT-domain. Popular among the spatial domain methods are the Projection onto Convex Sets (POCS) approach¹ and the Bayesian approach.² Most SR methods assume only camera motion between successive frames. Eren et al.³ extended POCS method to object-based high resolution in the spatial domain. The POCS approach was later extended to include DCT-domain quantization by Patti *et al.*,⁴ who obtained better results on compressed video, compared to spatial domain processing on decompressed video ignoring the DCT quantization. However, none of these works thoroughly investigate the effect of quantization of the pixel intensities (8-bits RGB or 8-bits YUV), nor the effects of inaccuracies in sub-pixel motion estimation (registration errors).

In this paper, we analyze the performance of the POCS-based SR reconstruction in both spatial and compressed domains. It is assumed that the low-resolution input video is captured with 8-bit/pixel accuracy in RGB domain (uncompressed case) or converted to 8-bit YUV representation (compressed case). Although it is well-known that SR reconstruction requires highly accurate sub-pixel motion estimation, the sensitivity of the resolution improvement to motion estimation accuracy is not analyzed in the literature. We demonstrate the needed accuracy with experimental results. Hence, this paper provides data for analysis of the performance of POCS-based SR reconstruction in practical applications.

In Section 2, we review the modeling of video acquisition, including the compression process. We also incorporate pixel intensity quantization in the previously established model. The SR reconstruction methods applied to uncompressed and compressed video are briefly summarized in Section 3. Section 4 presents the experimental conditions and results. Conclusions are provided in Section 5.

2. MODELING OF VIDEO ACQUISITION

In this section, we review modeling of low resolution (LR) video acquisition in spatial and compressed domains. As depicted in Figure 1, the video formation model¹ consists of several components: 1) Motion warping, which is due to the movement of the camera, and possible changes in the scene; 2) Sensor blurring, which is caused by integrating the received light over a finite sensor area; 3) Optical blurring caused by the imaging optics; 4) Motion blurring, due to the integration of signal during a nonzero aperture time; 5) Sampling on an arbitrary space-time lattice; 6) Sensor noise; 7) Pixel intensity quantization caused by digitization of analog data, e.g., 8 bits/color.

The input to the model is a spatially continuous reference image $f(x_1, x_2)$, and the output is the observed LR video, $g(m_1, m_2, k)$. By incorporating a motion model, the relation between the LR video and the desired SR reference image can be formulated as,

$$g(m_1, m_2, k) = \sum_{(n_1, n_2)} f(n_1, n_2) \cdot h(n_1, n_2; m_1, m_2, k) + v(m_1, m_2, k) \quad (1)$$

where $h(n_1, n_2; m_1, m_2, k)$ represents the linear shift-varying (LSV) blur function of the video acquisition process, $v(m_1, m_2, k)$ is the additive noise due to the LR sensor, and $f(n_1, n_2)$ denotes the discretized SR image. It is important to notice that Eqn. 1 provides a set of linear equations which relate the desired SR image to the LR video via a blur function $h(n_1, n_2; m_1, m_2, k)$, that needs to be computed very accurately for each LR pixel $g(m_1, m_2, k)$ as a function of the motion parameters.

The video formation model in Figure 1 was extended to include transform coding, as depicted in Figure 2.⁴ The motion compensated prediction $\hat{g}(m_1, m_2, k)$ is formed by motion compensating the previously decoded LR image or images with some motion vectors. Then, the residual signal $g_m(m_1, m_2, k)$ can be expressed in the following way:

$$g_m(m_1, m_2, k) = g(m_1, m_2, k) - \hat{g}(m_1, m_2, k) \quad (2)$$

If we incorporate Eqn. 1 into Eqn. 2, and then take 8×8 block-DCT of both sides, we get a new set of linear equations, which relate the DCT coefficients of the observed LR video $d(m_1, m_2, k)$ to the unknown SR image $f(n_1, n_2)$ through a motion model:

$$d(m_1, m_2, k) = \sum_{(n_1, n_2)} f(n_1, n_2) \cdot h_{dct}(n_1, n_2; m_1, m_2, k) - DCT_{op}(\hat{g}(m_1, m_2, k)) + DCT_{op}(v(m_1, m_2, k)) \quad (3)$$

where DCT_{op} denotes the operator for the 8×8 block-DCT. In this formulation, $h_{dct}(n_1, n_2; m_1, m_2, k)$ represents the new LSV blur function, which is built by applying a 8×8 block-DCT to the blur function $h(n_1, n_2; m_1, m_2, k)$. All the effects of motion modeling are incorporated into this function, and therefore it greatly depends on the actual motion in the video. Due the latter fact, performance of the SR reconstruction algorithm is strongly connected to the accuracy of motion estimation among the LR video frames.

Figure 1. Video acquisition model

Eqn. 3 is the main equation for the SR reconstruction process in the compressed domain. As depicted in Figure 2, after transform coding, the LR image DCT coefficients are quantized using the proper quantization step sizes that are determined by the picture and macroblock type. However, Eqn. 3 is not adequate to find the desired SR image by itself, since the exact DCT coefficients are unknown at the decoder side. Nevertheless, using the quantization information embedded in the MPEG bit stream,⁵ valid intervals for the quantized DCT coefficients can be rebuilt. This fact makes it possible to solve the SR reconstruction problem for the compressed domain in a POCS framework. The next section explains how the POCS method is used to reconstruct an SR image from both compressed and uncompressed LR images.

3. THE RECONSTRUCTION METHOD

In order to find a solution to the SR reconstruction problem, the linear set of equations provided by Eqn. 1 and Eqn. 3 are used appropriately in a POCS framework⁶ for the uncompressed and compressed LR images respectively. In the following subsections, we review how the POCS method is applied to these two cases.

3.1. Spatial-domain POCS

The method of POCS requires closed convex constraint sets to be defined within a well-defined space, such that the feasible solution is inside the intersection of these sets. When the goal is to reconstruct the SR image from uncompressed LR images, one constraint set $C(m_1, m_2, k)$ is defined for each pixel in the LR video in the following way:

$$C(m_1, m_2, k) = \{y(n_1, n_2) : |r^{(y)}(m_1, m_2, k)| \leq \delta_o(m_1, m_2, k)\} \quad (4)$$

where

$$r^{(y)}(m_1, m_2, k) = g(m_1, m_2, k) - \sum_{(n_1, n_2)} y(n_1, n_2) \cdot h(n_1, n_2; m_1, m_2, k) \quad (5)$$

is the residual associated with the SR estimate $y(n_1, n_2)$, and δ_o is the bound determined by the noise statistics.

Associated with each constraint set, there exists a projection operator $P(m_1, m_2, k)$, and it is defined as,

$$P(m_1, m_2, k)[y(n_1, n_2)] = y(n_1, n_2) + \begin{cases} \frac{(r^{(y)}(\cdot) - \delta_o(\cdot))h(n_1, n_2; \cdot)}{\sum_{o_1} \sum_{o_2} h^2(o_1, o_2; \cdot)}, & r^{(y)}(\cdot) > \delta_o(\cdot) \\ 0, & -\delta_o(\cdot) \leq r^{(y)}(\cdot) \leq \delta_o(\cdot) \\ \frac{(r^{(y)}(\cdot) + \delta_o(\cdot))h(n_1, n_2; \cdot)}{\sum_{o_1} \sum_{o_2} h^2(o_1, o_2; \cdot)}, & r^{(y)}(\cdot) < -\delta_o(\cdot) \end{cases} \quad (6)$$

where the ‘.’ function argument is used for m_1, m_2, k .

Using this operator, an initial estimate of the SR image is projected onto the constraint sets repeatedly until an estimation is found in the intersection of all constraint sets. In practice, however, the iterations are usually stopped when the visual quality of the SR image is acceptable, or the change in the MSE is below some threshold.

Figure 2. Overall video acquisition model with MPEG compression

3.2. DCT-domain POCS

If the LR input video is in compressed format, such as MPEG, H.263 or motion JPEG, the closed convex constraint sets $C(m_1, m_2, k)$ are defined⁷ at every point (m_1, m_2, k) as given in Eqn. 7.

$$C(m_1, m_2, k) = \left\{ y(n_1, n_2) : \left[\sum_{(n_1, n_2)} y(n_1, n_2) \cdot h_{dct}(n_1, n_2; m_1, m_2, k) - DCT_{op}(\hat{g}(m_1, m_2, k)) \right] \right. \\ \left. \in [b_l(m_1, m_2, k), b_u(m_1, m_2, k)] \right\} \quad (7)$$

where $b_l(m_1, m_2, k)$ and $b_u(m_1, m_2, k)$ are the quantization bounds corresponding to the quantized DCT coefficient at (m_1, m_2, k) , derived from the MPEG bit stream. The projection operator $P(m_1, m_2, k)$ which takes the current SR estimate $\hat{f}(n_1, n_2)$, and projects it onto the constraint sets can be written in the following way:

$$\hat{f} \bullet h_{dct} \equiv \sum_{(n_1, n_2)} \hat{f}(n_1, n_2) h_{dct}(n_1, n_2; \cdot), \quad \|h_{dct}\|^2 \equiv \sum_{(n_1, n_2)} h_{dct}^2(n_1, n_2; \cdot), \quad \hat{G} = DCT_{op}(\hat{g}(\cdot)) \quad (8)$$

$$P(m_1, m_2, k)[\hat{f}(n_1, n_2)] = \hat{f}(n_1, n_2) + \begin{cases} \frac{(b_l(\cdot) + \hat{G} - \hat{f} \bullet h_{dct}) h_{dct}(n_1, n_2; \cdot)}{\|h_{dct}\|^2}, & \hat{f} \bullet h_{dct} < b_l(\cdot) + \hat{G} \\ \frac{(b_u(\cdot) + \hat{G} - \hat{f} \bullet h_{dct}) h_{dct}(n_1, n_2; \cdot)}{\|h_{dct}\|^2}, & \hat{f} \bullet h_{dct} > b_u(\cdot) + \hat{G} \\ 0, & elsewhere \end{cases} \quad (9)$$

where the ‘ $\cdot$ ’ function argument is used for m_1, m_2, k .

Projection onto a constraint set $C(m_1, m_2, k)$ is done only if DCT coefficient is available at pixel (m_1, m_2, k) , and Eqn. 3 is valid through an accurate estimation of the LSV blur function, $h_{dct}(n_1, n_2; m_1, m_2, k)$.

4. EXPERIMENTS

We have run several experiments to analyze the factors that may affect the performance of the SR algorithm. The POCS framework that we use for the SR reconstruction already gives us a few hints about those factors, such as the size of the constraint sets defines the existence of a solution and also affects the convergence rate. If the constraint set size is not chosen to be larger than the possible amount of noise in the image, then the SNR's of the SR estimates start to drop after a few iterations. However, constraint set size is not the only factor causing SNR's to drop. Any slight deviation from the actual camera model may set the solution far away from the correct solution set. Therefore, the blurring operators, e.g., optical blur, should be consistent with the original models as much as possible.

For the experiments, a number of LR images are created using the camera model given in Section 2. The optical blur function in the model is set to a gaussian filter with a variance of 0.8 and a 3×3 support on the SR grid, and the SR sensor element is assumed to have a square support with uniform response. The input is a gray-level image of size 352×288 , and the LR images created have half size of the input signal in both vertical and horizontal directions. Since some of the experiments are done in the DCT-domain, the simulated LR images are then compressed using the MPEG4 Simple Profile at different bit rates. Eight of the created LR images are used to reconstruct the input image, which is assumed to be the original SR image. In the following subsections, we explain the factors that limit the performance of the SR reconstruction, by describing the corresponding experimental set-up in detail and presenting the experimental results. For all the experiments, the PSNR results of the SR estimates are given up to fifteen iterations since this number of iterations was found to be adequate to describe the trend of the iterative process.

4.1. Interpolation error

In the POCS-based SR reconstruction algorithms, the first SR estimate is usually a bi-linearly interpolated reference LR image. The quality of the first SR estimate directly affects the visual and SNR quality of the resulting SR image, which is built by a number of iterations. To measure the effect of the interpolation filter on the SR reconstruction algorithm, we used an optimal filter, the Wiener filter,⁸ to build up the first SR estimate, and then ran the SR algorithm. During the experimentation, the other factors that might affect the quality of the SR image were turned off, so there was no pixel-domain quantization error, original motion vectors (the ones used in creating the LR images) were used, and the low-resolution images were uncompressed. Regarding the camera parameters, the optical blur function was set to the Gaussian filter that was used for creating the LR images.

Figures 3-a,b show the PSNR results of the SR estimate with respect to the original SR image using Wiener filter of different lengths, i.e., 2-tap, 6-tap, and 8-tap. 2-tap Wiener filter corresponds to the bi-linear interpolation. Because the improvement obtained by using a longer filter than the 8-tap is negligible, the results are given up to 8-tap Wiener filter. As seen from the figures, 6-tap and 8-tap filters result in very close PSNRs and provide an improvement of about 1-2 dB over the 2-tap filter. Therefore, the Wiener filter is strongly advised in practical applications of SR reconstruction, because it provides a substantial improvement over bi-linear interpolation for the first SR estimate without added cost.

Since we created the low-resolution images, we had the opportunity to make them noise free. This enables us to choose δ_o freely as long as it is different from zero. To observe the effect of different δ_o values, we ran the SR algorithm using $\delta_o = 0.001$ and $\delta_o = 0.1$; these results are given in Figure 3-a and Figure 3-b respectively. As observed from Figure 3-a, a smaller δ_o results in a higher PSNR for the SR estimates. This conclusion is consistent with the theory of POCS because the constraint set size is determined by δ_o , and as the set size decreases, the possibility of reaching the true solution set increases as long as there is no noise in the LR images. Therefore, δ_o should be chosen to be larger than the noise in the image to ensure a solution exists, but small as possible to achieve higher PSNR for the SR estimates.

4.2. Pixel-domain quantization error

For consumer oriented cameras, the signals from the single-chip color sensor are converted by applying 10 bits/color linear quantization. Afterwards, the color sensor data is processed to produce RGB signals, which are then gamma-protected to produce what is called “sRGB” image data. This gamma-corrected image data is then converted to 8-bit YUV type signals, and JPEG compressed. We call the integer truncation that takes place inside the camera during video acquisition “pixel-domain quantization error”. Although we started with an assumption of a 3-sensor camera model and did not have to deal with the color filter array (CFA), we did simulate the noise due to a decrease in the bit precision, from 10 bit/color to 8 bit/channel, by a simple nearest integer truncation.

In the experiments, two kinds of uncompressed LR images were used as the input data, the ones that were saved as floating data type, and the ones that were rounded and saved as integer data type. The latter ones included pixel-domain quantization error. For the SR reconstruction, the first SR estimate was built by applying a 6-tap Wiener filter to the reference LR image, and δ_o was set to 0.1. Original motion vectors were used to warp the created LR images onto the reference image. The optical blur function in the camera model was set to the one that was used for creating the LR images.

In Figure 4, we compare the PSNRs of SR estimates to show the effect of pixel-domain quantization error. From the experimental results, it is observed that although the pixel-domain quantization is the only limiting factor for the SR reconstruction, the PSNRs of the SR estimates start to decrease after some point, and continue to decrease as the iterations proceed. This is the correct expectation from a POCS-based algorithm, because the constraint set size was set to a value (0.1) smaller than the possible amount of noise in the LR images, which was limited by the rounding error (0.5). This again shows the importance of matching δ_o to the noise in the image.

4.3. Motion estimation error

The accuracy of sub-pixel motion vectors plays a crucial role in the performance of the SR reconstruction, as explained in Section 2. To see the effect of sub-pixel motion accuracy, we computed the motion field among the LR images using the MPEG7 global motion estimator. In the experiments, uncompressed LR images were used, and since they were built by rounding to the nearest integer, they included the pixel-domain quantization error. The first SR estimate was built by applying a 6-tap Wiener filter to the reference LR image. The optical blur function was set to the gaussian filter used in the creation of the LR images.

Figure 4 shows the PSNR results of the SR estimates for $\delta_o = 0.1$. From the experimental results, the decrease in the SNR's is clearly seen when the estimated motion vectors are used in the SR reconstruction. Since the estimated motion field has some possible amount of noise, there is a deviation from the actual camera model, and this causes the PSNRs of the SR estimates to decrease as the iterations proceed. As observed from Figure 4, the maximum PSNR decreases by about 2.7 dB when estimated motion field is used. This fact makes the motion estimation error the most limiting factor on the performance of the SR reconstruction. Therefore very robust motion estimation algorithms are required for the SR reconstruction in practical applications.

4.4. Optical blur error

The performance of the SR reconstruction also depends on the correct choice of the parameters for the optical blur function. The original blur function that is used in the creation of the LR images is a gaussian filter of variance 0.8. To observe the effect of different variances, we performed experiments using variances 0.7, 0.6 and 1.0 for the gaussian filter with the same 3×3 support on the SR grid.

During experimentation, uncompressed LR images were used. There was no pixel-domain or DCT-domain quantization. The first SR estimate was built by applying a 6-tap Wiener filter to the reference LR image. No motion estimation was performed, and original motion vectors were used for motion warping. Figure 5 shows the PSNR results of the SR estimates for $\delta_o = 0.1$. As observed from the results, the incorrect filter parameter results in a decrease in the PSNRs of the SR estimates. This conclusion is consistent with the theory of POCS because there is a deviation from the actual camera model, whether the variance is smaller or larger than the original one. However, the maximum decrease in the PSNRs of the SR estimates is about 1.75 dB. Therefore we can conclude that the optical blur error does not affect the performance of the SR reconstruction as much as the motion estimation error.

4.5. DCT-domain quantization error

Experiments were performed to measure the effect of the DCT-domain quantization onto the SR reconstruction of compressed LR images. For the experiments, created spatial-domain LR images were compressed using the MoMuSys MPEG4 codec with H.263 type quantization. In the DCT-domain SR reconstruction, quantization bounds form the constraint sets of the POCS algorithm. Therefore, the quantization bounds were derived from the MPEG4 bit stream to build the convex constraint sets. The experiments were repeated at different bitrates (14kb/fr, 10kb/fr, 6kb/fr, 3kb/fr) using corresponding quantization scalers (2, 3, 5, 10). As the optical blur function, the one used for creating the LR images was implemented, and the first SR estimate was built by applying a 6-tap Wiener filter to the reference LR image.

To measure the effect of accuracy of sub-pixel motion vectors (MV), we performed three experiments by using i) the original MVs used to create the LR images, ii) the MVs estimated at the encoder side using the uncompressed LR images, and iii) the MVs estimated at the decoder side using the decompressed LR images. The PSNR results for the three cases are shown in Figures 6-a,b,c respectively. It is observed from the results that as the quantization scaler (mquant) increases, the PSNRs of the SR estimates decrease. This is because a larger mquant means more compression, and less quality for the decompressed LR image. Since the first SR estimate in the SR reconstruction is the interpolation of the decompressed reference LR image, the first SR estimate has a smaller PSNR for a larger mquant. When the original motion field is used in the SR reconstruction, it is observed that the PSNR results for any mquant value converge because there is not any deviation from the actual camera model. When the estimated motion vectors are used in the SR reconstruction, the PSNRs of the SR estimates start to decrease after some point, since the camera model is not correct because of the incorrect motion field. As the error in the estimation of the motion field increases, the decrease in the PSNRs increases. This can be

Figure 3. PSNRs of SR estimates using Wiener filter of different lengths with (a) δ_o of 0.001 (b) δ_o of 0.1 .

observed from the comparison of Figure 6-b with Figure 6-c, where maximum PSNR decreases by about 0.67 dB when the mquant is 2. Motion estimation error affects the quality of the SR estimate more for smaller mquant values. This is because a small mquant means a small convex set, and as explained in Subsection 4.1, the size of the convex set affects the PSNR of the SR estimates directly.

5. CONCLUSION

In this paper, we studied the factors that affect the performance of a POCS-based SR reconstruction algorithm. For this purpose, we conducted controlled experiments in different scenarios. In addition, we modified the modeling stage of video acquisition by adding spatial-domain quantization and proposed the usage of a Wiener filter for the construction of the first SR estimate in the iterative POCS solution. The experimental results have been shown to be consistent with the theory of POCS. Our future work includes application of this algorithm to real-life scenes, such as dynamic scenes, to achieve high quality SR images, by making use of the knowledge of POCS's limitations as described in this paper.

ACKNOWLEDGMENTS

This research has been funded by the Center for Electronic Imaging Systems under a New York State STAR grant.

REFERENCES

1. A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time," *IEEE Trans. on Image Processing* **6**, pp. 1064–1075, 1997.
2. R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. on Image Processing* **5**, pp. 996–1011, 1996.
3. E. Eren, M. I. Sezan, and A. M. Tekalp, "Robust, object-based high-resolution image reconstruction from low-resolution video," *IEEE Trans. Image Proc.* **6**, pp. 1446–1451, 1997.

4. A. J. Patti and Y. Altunbasak, "Super-resolution image estimation for transform coded video with application to mpeg," *Proc. IEEE Int. Conf. Image Processing* **3**, pp. 179–183, 1999.
5. A. M. Tekalp, *Digital Video Processing*, Prentice-Hall, New Jersey, 1995.
6. M. I. Sezan, "An overview of convex projections theory and its application to image recovery problems," *Ultramicroscopy* **40**, pp. 55–67, 1992.
7. Y. Altunbasak, A. J. Patti, and R. M. Mersereau, "Super-resolution still and video reconstruction from mpeg-coded video," *IEEE Trans. on Circ. and Sys. for Video Tech.* **12**, pp. 217–226, 2002.
8. O. Werner, "Drift analysis and drift reduction for multiresolution hybrid video coding," *Signal Proc. Image Communication* **8**, pp. 387–409, 1996.

Figure 4. PSNRs of SR estimates with and without pixel-domain quantization using true and estimated MVs.

Figure 5. PSNRs of SR estimates with different variances for the gaussian filter.

Figure 6. PSNRs of SR estimates with (a) original MVs, (b) MVs estimated at the encoder side, (c) MVs estimated at the decoder side.