Perceptually-Adaptive Image Super-Resolution using Statistical Methods

ALEXANDER WONG
Department of Electrical and Computer Eng.
University of Waterloo
Waterloo, Ontario, N2L 3G1
CANADA
a28wong@engmail.uwaterloo.ca
http://www.eng.uwaterloo.ca/~a28wong

WILLIAM BISHOP
Department of Electrical and Computer Eng.
University of Waterloo
Waterloo, Ontario, N2L 3G1
CANADA
wdbishop@uwaterloo.ca
http://www.ece.uwaterloo.ca/~wdbishop

Abstract: - Multi-frame image super-resolution makes use of a set of low-resolution images to reconstruct one or more high-resolution images. This paper presents a novel super-resolution algorithm that uses perceptually important content characteristics such as edges, texture, and brightness to improve visual quality. The super-resolution algorithm introduces perceptually-adaptive constraint relaxation to optimize the image for the human vision system. Experimental results show that the super-resolution algorithm improves visual quality both quantitatively and qualitatively when compared with standard techniques.

Key-Words: - Perceptually-adaptive, image super-resolution, statistical estimation

1 Introduction

Image super-resolution techniques use low-resolution (LR) images to produce high-resolution (HR) images. In the case of multi-frame super-resolution, a set of LR images is used to reconstruct one or more HR images. Super-resolution techniques have gained popularity in various fields that make use of image processing and computer vision, such as digital photography, remote sensing, and security surveillance. This is due to the fact that the resolution achieved by imaging systems is often constrained by technical limitations and other factors such as cost. Super-resolution methods provide an effective alternative to more expensive and complex solutions.

Reconstruction of a single HR image from a set of LR images is only possible if the set of LR images exhibit informational differences. If all of the LR images are identical, no additional information can be extracted about the original HR image. By taking unique information from each LR image, it is possible to interpolate sub-pixel values and reconstruct a single HR image with informational content that cannot be provided from just one LR image.

A large number of super-resolution techniques have been proposed. These can be generally categorized as follows:

1. Single-frame super-resolution [1-2]
2. Multi-frame image super-resolution [3-12]

In *single-frame super-resolution* techniques, a single LR image is used to reconstruct a high-resolution image. Such techniques yield a higher-quality HR image that possesses no more information than the base LR image. Such techniques are primarily useful for image enhancement purposes. Of greater interest in recent years are *multi-frame image super-resolution* techniques, where a number of low-resolution images are used to reconstruct one or more HR images that possess more information detail than any one LR image.

One prominent class of multi-frame image super-resolution techniques is based on Projection on Convex Sets (POCS). First presented by Sauer and Allebach [3], the basic premise behind POCS-based image super-resolution is that a solution to a problem can be obtained by projecting it onto each constraint in a convex constraint set. In the case of the image super-resolution problem, the estimated solution is the HR image and the constraint set is defined to ensure data consistency with the LR images. From an initial guess based on the first LR image, the estimated solution can be repeatedly updated by projecting it onto the constraint set defined by subsequent LR images.

Another popular technique for multi-frame image super-resolution is the iterative back-projection method proposed by Irani and Peleg [4]. In this algorithm, an initial guess of the HR image is back-projected to form projected LR images by applying degradation models on the HR image. The error between the projected images and the original LR images is calculated and the estimated HR solution is updated based on the error. Elad and Feuer [5] propose a hybrid approach that combines the quadratic constraint in a maximum likelihood estimator with convex constraints. This new convex optimization problem can then be solved using efficient iterative least squares solvers. The main benefit to this algorithm over POCS is that there is a single optimal solution. Researchers have made use of conjugate gradient methods [6-7] to develop efficient image super-resolution methods. Other multi-frame image super-resolution techniques include Bayesian methods [8], Markov random fields [9], and techniques based on machine learning [10-11].

Recent research in multi-frame image super-resolution has focused on improving computational efficiency. However, little research has been conducted on improving the visual quality of the resultant HR image. Existing adaptive super-resolution algorithms attempt to improve image quality [13] but little consideration is given to the characteristics of the human vision system. The perception of the human vision system is particularly important in digital photography, where the primary goal is to improve the visual quality of the image as it is perceived by an individual. Therefore, it is desirable for a multi-frame image super-resolution algorithm to adapt based on perceptually important characteristics of the LR images used to generate the HR image.

The main contribution of this paper is an adaptive multi-frame image super-resolution algorithm that utilizes perceptually important characteristics such as edges, texture, and brightness to improve visual quality. In this paper, the theory underlying the proposed algorithm is described and explained in detail in Section 2. An outline of the algorithm is provided in Section 3. The testing methods and test data are outlined in Section 4. Experimental results demonstrating the effectiveness of the proposed super-resolution algorithm compared to the use of uniform prior constraint models are presented in Section 5. Finally, conclusions are drawn in Section 6.

2 Theory

This section introduces the theory behind the proposed perceptually-adaptive multi-frame image super-resolution algorithm. For the remainder of this paper, the term *super-resolution* is used to refer to multi-frame image super-resolution. First, the theory behind the super-resolution problem formulated as an input estimation problem is reviewed. Second, a set of perceptually important characteristics of images are described in detail. Finally, the theory behind adaptive constraint relaxation is explained.

2.1 Super-Resolution Problem

Before describing the proposed algorithm, it is necessary to define the super-resolution problem. A single LR image can be viewed as a single HR image that has been degraded due to motion, imaging conditions, and down-sampling. A simple imaging model can be defined as

$$I_{\mathrm{LR}} = HI_{\mathrm{HR}} + e \qquad (1)$$

where H is an overall operator for the degradation caused during the imaging process (such as blurring, decimation, warping, motion, and etc.), I_{HR} and I_{LR} are the HR and LR images respectively, and e represents the additive noise. The overall degradation operator can be expressed as a combination of different degradation operators. For example, a typical model used to represent overall degradation for the super-resolution problem is defined as being a combination of decimation (D) and blur (B) operators. Therefore, (1) can be expressed as the following:

$$I_{\mathrm{LR}} = \left(DB\right) I_{\mathrm{HR}} + e \qquad (2)$$

For multi-frame super-resolution, the set of LR images $\{I_{\mathrm{LR},1}, I_{\mathrm{LR},2}, \ldots ,I_{\mathrm{LR},n}\}$ used to reconstruct the HR image are viewed as being derived from a single HR image I_{HR}. Motion models must be taken into account for each LR image to align properly in the reference coordinate system. Thus, the imaging model for each LR image can be expressed as

$$I_{\mathrm{LR},k} = \left(D_k B_k M_k\right) I_{\mathrm{HR}} + e_k,\ 1 \le k \le n \qquad (3)$$

where $I_{\mathrm{LR},k}$ is the k^{th} LR image and M_k is the motion operator for the k^{th} image relative to a reference image, and D_k and B_k are the decimation and blur operators for the k^{th} image respectively. Hence, if the degradation operators are known, the super-resolution problem can be defined as an input estimation problem, where a single HR image (I_{HR}) is estimated from information obtained from the set of n LR images $\{I_{LR,1}, I_{LR,2}, \ldots, I_{LR,n}\}$. Thus, the super-resolution problem can be solved using statistical methods.

The super-resolution problem as defined above can be visualized using a simple example, as illustrated in Fig.1. For the purpose of this example, assume that two $m \times m$ images $I_{LR,1}$ and $I_{LR,2}$ are acquired for a particular scene at slightly different positions. Let each pixel value represent a measurement from the scene. To obtain a $2m \times 2m$ image I_{HR}, the images are aligned with each other and the coordinate of each measurement in each image is then aligned and up-scaled by a factor of two in each dimension. It can be seen that each $m \times m$ image possesses unique measurements that can be used to provide a better estimate than what either of

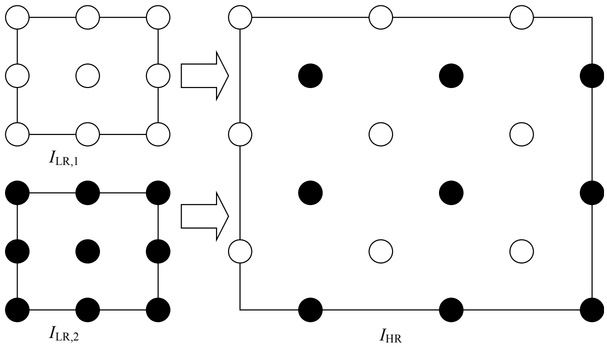

Fig.1. Visualization of super-resolution problem

the images can provide alone.

2.2 Perceptually Important Characteristics

Super-resolution techniques have not traditionally taken into account the perception of the human vision system. By exploiting the characteristics that are perceptually important to the human vision system, it is possible to improve the overall visual quality of an image. The following perceptually important content characteristics are exploited in the proposed super-resolution algorithm:

1. Edges
2. Texture activity
3. Brightness

These characteristics can be extracted from an interpolated version of a reference LR image chosen from the set of LR images.

2.2.1 Edges

To determine the edges of an image, an edge detection algorithm is applied to the reference LR image to create a binary edge map (E) with edge and non-edge pixels being represented by 1 and 0 respectively. An edge rating (ER) for a pixel is calculated based on its neighboring pixels:

$$ER(x, y) = \max(E_{5\times 5 \text{ neighborhood}}(x, y)) \tag{4}$$

The resulting edge rating (ER) is a 1 if an edge pixel is found in any of the neighboring pixels or in the current location.

Edges are crucial to the way the human vision system perceives the environment. As a result, the human vision system is very sensitive to edge degradation in an image. Furthermore, the human vision system is less sensitive to noise in edge regions than smooth regions. Therefore, it is important that the pixels in edge regions of the HR image closely resemble the original information obtained from the LR images to preserve edges for better visual quality.

2.2.2 Texture Activity

To determine the texture activity of a pixel, a texture rating (TR) is calculated based on its neighboring pixels:

$$TR(x, y) = s^2_{7\times 7 \text{ neighborhood}}(x, y) \tag{5}$$

where $s^2_{7\times 7 \text{ neighborhood}}(x,y)$ is the spatial variance of pixel intensities in a 7×7 neighborhood. The size of the neighborhood chosen for texture activity is larger than the one chosen for edges to ensure an adequate representation of the local texture activity. The human vision system is less sensitive to noise in regions with high texture activity. Therefore, the pixels in these regions of the HR image can closely resemble the original information obtained from the LR images to preserve fine texture details without introducing highly perceivable noise. The spatial variance is used as a simple measurement of texture activity. This measurement has a low computational complexity and delivers a good measurement of texture activity.

2.2.3 Brightness

Finally, the overall brightness rating (BR) is determined based on its neighboring pixels:

$$BR(x, y) = \mu_{7\times 7 \text{ neighborhood}}(x, y) \qquad (6)$$

where $\mu_{7x7 \text{ neighborhood}}(x,y)$ is the sample mean of the pixel intensities in a 7×7 neighborhood. The size of the neighborhood chosen for brightness is the same as the one chosen for texture activity. The human vision system is less sensitive to noise in dark areas. Therefore, the pixels in dark regions of the HR image can closely resemble the original information obtained from the LR images to preserve fine details without introducing highly perceivable noise. The sample mean is used as a measurement of brightness because of its low computational complexity.

2.3 Adaptive Constraint Relaxation

The super-resolution problem posed in (3) can be expressed in matrix-vector form as the following:

$$\begin{bmatrix} (\underline{I}_{\mathrm{LR},1})_: \\ (\underline{I}_{\mathrm{LR},2})_: \\ \vdots \\ (\underline{I}_{\mathrm{LR},k})_: \\ \vdots \\ (\underline{I}_{\mathrm{LR},n})_: \end{bmatrix} = \begin{bmatrix} D_1 B_1 M_1 \\ D_2 B_2 M_2 \\ \vdots \\ D_k B_k M_k \\ \vdots \\ D_n B_n M_n \end{bmatrix} (\underline{I}_{\mathrm{HR}})_: + \begin{bmatrix} \underline{e}_1 \\ \underline{e}_2 \\ \vdots \\ \underline{e}_k \\ \vdots \\ \underline{e}_n \end{bmatrix} \qquad (7)$$

where $(\underline{I}_{\mathrm{LR},k})_:$ is the k^{th} LR image and $(\underline{I}_{\mathrm{HR}})_:$ is the HR image lexicographically ordered. This can be expressed in a simpler form as the following:

$$\overline{I}_{LR} = H(\underline{I}_{HR}) + \overline{\underline{e}} \qquad (8)$$

This can then be solved using a linear systems solver. One of the difficulties with the super-resolution problem posed in (8) is that the problem is an ill-posed inverse problem. Practical super-resolution scenarios typically result in underdetermined systems, where there exist an infinite number of solutions and uniqueness fails. Therefore, it is necessary to introduce a prior model for the HR image to make the problem well-posed. The regularized optimization problem can be expressed as the following:

$$\hat{\underline{I}}_{HR} = \mathrm{ArgMin}\left\{ \left\| \overline{I}_{LR} - H(\hat{\underline{I}}_{HR}) \right\| + \lambda \left\| L\hat{\underline{I}}_{HR} \right\| \right\} \qquad (9)$$

where $\hat{\underline{I}}_{\mathrm{HR}}$ is the estimated HR image, L is the prior constraints matrix and λ is the relaxation value that controls the degree of approximation of the system. A large value of λ results in a problem that is a poor approximation to the original problem but is very well conditioned. A small value of λ results in a problem that is faithful to the original problem but is poorly conditioned. In other words, a large value results in a smoother estimation and a smaller value results in a closer resemblance to the original information obtained from the LR images. Methods such as generalized cross-validation (GCV) can be used to select a global value for λ. However, given the diverse nature of content within an image and the nature of the human vision system, it is difficult to select a global value of λ that will satisfy all situations in terms of visual quality. One approach to this problem is to adaptively adjust the relaxation value based on the perceptually important characteristics described in Section 2.2. If a pixel lies on an edge or surrounded by edges, then a low relaxation value is used to preserve the edge detail. If a pixel lies on a dark region or a region with high texture activity, then a low relaxation value should also be used to preserve fine details without introducing highly perceivable noise. For smooth regions, a high relaxation value is used to provide a smoother estimation as the human vision system is highly sensitive to noise in such regions.

3 Algorithm Outline

Based on the theory presented, the proposed super-resolution algorithm can be summarized as follows:

Given n LR images with one designated as the reference image:

1. Calculate the edge rating ER, texture rating TR, and brightness rating BR for each pixel from the interpolated version of the reference image.
2. For each pixel:
    a. If $ER(x,y)$=1, the relaxation value is $\lambda(x,y) = \lambda_{\text{low}}$
    b. If $BR(x,y) < T_{\text{brightness}}$ or $TR(x,y) < T_{\text{texture}}$, the relaxation value is $\lambda(x,y) = \lambda_{\text{low}}$
    c. Otherwise, $\lambda(x,y) = \lambda_{\text{high}}$
3. Solve the regularized optimization problem

posed in (9) to obtain the final HR image.

4 Testing Methods

To test the effectiveness of the proposed adaptive super-resolution method, three test sets consisting of sixteen 8-bit grayscale LR images were created as described below:

- **LENA:** Set of 128×128 LR images generated from 512×512 HR image of Lena.
- **IKONOS:** Set of 64×64 LR images generated from a 256×256 HR segment of an IKONOS image of Athens Olympics Sports Complex, Greece.
- **CARDS:** Set of 32×32 LR images generated from a 128×128 HR image showing playing cards.

Each LR image was synthetically generated from the original HR image by undergoing global translation motion, 4×4 uniform blur, and decimation by a factor of 4 for each dimension. The proposed super-resolution algorithm was then used to reconstruct an estimated HR image at the same resolution as the original image, using 9 of the 16 LR images. The LSQR algorithm [13] was performed for a maximum of 30 iterations to solve the optimization problem. The bilinear interpolated version of the reference LR image was used as an initial guess for the HR image. For the proposed method of relaxing prior constraints, a second-order thin-plate smoothness constraint was used and λ_{high} was set to 1 and λ_{low} was set to 0.1. The standard method of using a uniform prior model is provided for comparison. The uniform prior model utilizes a second-order thin-plate smoothness constraint model and $\lambda = 1$. For a quantitative comparison, the PSNR of the resultant images are provided.

5 Experimental Results

The quantitative results are shown in Table 1. The quantitative measurements show PSNR gains in the images produced using the proposed super-resolution algorithm when compared to the images produced using the uniform prior model. All test cases demonstrate some quantitative improvement ranging from +0.39 dB to +1.23 dB. The resultant images for CARDS and IKONOS are shown in Fig. 2 and Fig.3 respectively. It can be observed that the estimated HR images produced with adaptive

(a) Original HR image (b) Bilinear interpolation of a reference LR image

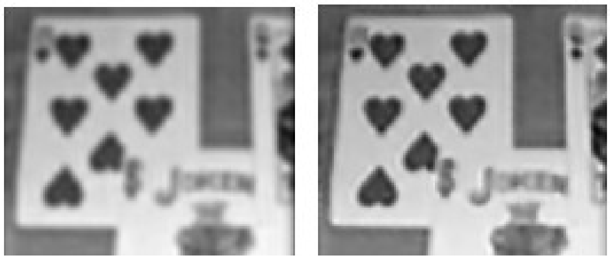

TABLE 1
IMPROVEMENTS IN PSNR FROM SUPER-RESOLUTION

Image	PSNR (dB)		PSNR Gain using Proposed Algorithm (dB)
	Uniform Prior Constraints	Proposed Algorithm	
LENA	28.19	28.58	+0.39
IKONOS	21.64	22.87	+1.23
CARDS	22.01	22.91	+0.90

constraint relaxation preserves edge details noticeably better than those produced with the uniform prior model. Furthermore, the proposed algorithm retains smoothness and reduces perceptible noise in the constant regions. This is particularly the case for the CARDS test, where the letters and numbers on the cards are noticeably more readable. These results demonstrate the effectiveness of the proposed algorithm for improving visual quality.

6 Conclusions

This paper presents a perceptually-adaptive image super-resolution algorithm based on adaptive constraint relaxation. Experimental results demonstrate the effectiveness of the proposed super-resolution scheme both quantitatively and qualitatively over standard techniques. It is believed that this method can be used successfully for improving visual quality in areas such as digital photography.

References:

[1] A. Patti and Y. Altunbasak, "Artifact reduction for set theoretic super resolution image reconstructions with edge adaptive constraints and higher-order interpolands," in IEEE Trans. Image Processing, 1999, Vol. 10, No. 1, pp. 179-186.
[2] W. Freeman, T. Jones, E. Pasztor, "Example-based super-resolution," in IEEE Computer Graphics and Applications, 2002, Vol. 1, No. 2, pp. 56-65.
[3] K. Sauer and J. Allebach, "Iterative reconstruction of band-limited images from nonuniformly spaced samples," in IEEE Transactions on Circuits and Systems, 1987, Vol. 34, No. 12, pp. 1497-1506.
[4] M. Irani and S. Peleg, "Improving resolution by image registration," in GMIP, 1991, Vol. 53, pp. 231–239.
[5] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," in IEEE Trans. Image Processing, 1997, Vol. 6, No.12, pp. 1646–1658.
[6] A. Zomet and S. Peleg, "Efficient super-resolution and applications to mosaics," in ICPR, 2000, Vol. 1, pp. 579–583.
[7] N. Nguyen, P. Milanfar, and G. Golub, "A Computationally Efficient Superresolution Image Reconstruction Algorithm," in IEEE Transactions on Image Processing, 2001, Vol. 10, No. 4, pp. 573-583.
[8] F. Cortijo, S. Villena, R. Molina, and A. Katsaggelos, "Bayesian super-resolution of text image sequences from low resolution observations," in Proc. Seventh International Symposium on Signal Processing and Its Applications, 2003, Vol. 1, pp. 421 - 424.
[9] D. Rajan and S. Chaudhuri, "Data fusion techniques for super-resolution imaging," in Information Fusion, 2002, Vol. 3, pp. 25-38.
[10] D. Capel and A. Zisserman, "Super-resolution from multiple views using learnt image models," in IEEE Computer Vision and Pattern Recognition, 2001, Vol. 2, pp. 627-634.
[11] C. Miravet and F. Rodriguez, "A hybrid MLP-PNN architecture for fast image superresolution," in Lecture Notes in Computer Science, 2003, Vol. 2714, pp. 401-408.
[12] T. Ono, H. Hasegawa, I. Yamada, K. Sakaniwa, "An adaptive super-resolution of videos with noise information on camera systems," in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, March 2005, Vol. 2, pp. 857-860.
[13] C. Paige and M. Saunders, “LSQR: an algorithm for sparse linear equations and sparse least squares,” in ACM Transactions On Mathematical Software, 1982, Vol. 8, No. 1, pp. 43-71.

(a) Original HR image

(b) Bilinear interpolation of a reference LR image

(c) Super-resolution using uniform prior constraints

(d) Super-resolution using proposed algorithm

Fig.3. Multi-frame super-resolution of IKONOS