14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

HIGH-QUALITY IMAGE INTERPOLATION BASED ON MULTIPLICATIVE SKELETON-TEXTURE SEPARATION

Takahiro Saito, Yuki Ishii, Yousuke Nakagawa, and Takashi Komatsu

Dept. of Electronics and Informatics Frontiers, High-Tech Research Center, Kanagawa University
3-27-1 Rokkakubashi, 221-8686, Yokohama, Japan
phone: +(81)45 481 5661, fax: +(81)45 491 7915, email: saitot01@kanagawa-u.ac.jp
web: www.ku-hrc.jp

ABSTRACT

This paper presents a high-quality interpolation approach that can adjust edge sharpness and texture intensity to reconstruct an image according to user's taste in picture quality. Our interpolation approach first resolves an input image I into its skeleton image U and its texture generator V and its residual image D such that I = U·V + D, and then interpolates each of the three components independently with a proper interpolation method suitable to each. The skeleton image is a bounded-variation function meaning a cartoon approximation of I, and interpolated with a super-resolution deblurring-oversampling method that interpolates sharp edges without producing ringing artifacts. The texture generator is an oscillatory function representing regular distinct textures, and interpolated with a standard linear interpolation algorithm. The residual image is a function representing irregular weak textures and reconstruction errors, and interpolated with a statistical re-sampling interpolation algorithm.

1. INTRODUCTION

Many image interpolation methods [1]~[8] have been proposed up to now, but each of them suffers from its peculiar artifacts. The most popular linear interpolation methods undergo aliasing distortions, image blurring and jaggy artifacts. In the linear interpolation methods, averaging across edges is usually performed, and it blurs sharp edges. To avoid image blurs, sharpening should be introduced into the interpolation, but it produces ringing artifacts near edges. Many algorithms have been proposed to improve picture quality of interpolated images. One approach to this problem is to prevent interpolation from crossing edges, and along this line most of nonlinear interpolation methods try to interpolate images along an edge direction [2]~[4]. These edge-directed interpolation methods usually reconstruct a sharper interpolated image, but often suffer from severe visual degradation in fine textured image areas and provide disordered bumpy visual impression. Iterative methods such as PDE-based schemes [5]~[7] and projection onto convex sets (POCS) schemes [8], constrain continuity of sharp edges and find the appropriate solution through iterations. These iterative methods reproduce sharp edges without producing ringing artifacts, but they tend to weaken fines textures.

In this paper, by making it possible to use interpolation algorithms suitable to different image ingredients simultaneously, we construct a high-quality image interpolation approach that can adjust edge sharpness and texture intensity to reconstruct an image according to user's taste in picture quality. Our image interpolation approach over-samples a skeleton and a texture and a residual of each image independently. Vese and Osher proposed an algorithm to resolve a given input image I into its skeleton image U and its texture image V such that $I \approx U + V$ [9]. However, it does not necessarily achieve satisfactory separation for really observed images [10] in that the separated factors U, V are correlated to each other. In this paper, instead of this additive skeleton-texture separation model, we employ a separation algorithm based on a multiplicative skeleton-texture separation model to apply it to the interpolation problem. Our interpolation approach first resolves an image I to be interpolated into its skeleton image U and its texture generator V and its residual image D such that $I = U \cdot V + D$, and then interpolates each of the three components independently with a proper interpolation method suitable to each component. Since the skeleton image U means a cartoon approximation of the image I and lives in the bounded-variation function space, its proper interpolation method is a super-resolution deblurring-oversampling method [5] that interpolates sharp edges without producing ringing artifacts. The texture generator V is an oscillatory function representing regular clear textures, and its proper interpolation method is a standard linear interpolation method. The residual image D is a function representing irregular faint textures and reconstruction errors, and its proper interpolation method is a statistical re-sampling interpolation method.

2. SKELETON-TEXTURE SEPARATION

2.1 Image Formation Model

A simple geometrical-optical model for image formation is a result of viewing Lambertian non-flat surface patches [11]. According to the model, an observed image may be considered as the projection of the real 3-D world surface normal **N** onto the light source direction **L**, multiplied by the albedo $r(x,y)$. The albedo r captures the characteristics of the 3-D object's material. The brightness of the observed image at each pixel location (x,y) may be represented by

$$I(x,y) = r(x,y) \cdot \left(\mathbf{N}^T(x,y) \cdot \mathbf{L}\right). \quad (1)$$

This means that the image brightness I captures the change in material via the albedo r that multiplies the normalized shading image I_s,

$$I_S(x,y) = \mathbf{N}^T(x,y) \cdot \mathbf{L}. \quad (2)$$

The simplest image-formation model assumes that the albedo r is constant within a given object in the image. Such a simple image-formation model does not account for image textures.

2.2 Texture Models for Skeleton-Texture Separation

To account for image textures, we need to extend the geometrical-optical image-formation model. There are two possible factors producing textures on images: perturbations to the surface normal and perturbations to the object's albedo. On the other hand, scene illu-

mination varies spatially more smoothly, and the variation makes little contribution to image textures.

2.2.1 Additive Model and Skeleton-Texture Separation

In the case of the perturbed surface normal, like the bump mapping technique in CG, assuming that the surface normal **N** within each non-flat surface patch is written as

$$\mathbf{N}(x,y) = \left(\mathbf{I} + \mathbf{P}(x,y)\right) \cdot \mathbf{N}_S(x,y), \tag{3}$$

where the matrix **I**+**P** is a rotation matrix and **P** is a perturbation matrix changing around the null matrix **0** and **I** is the identity matrix and $\mathbf{N}_S$ is the average surface normal within the surface patch, then the additive model for the skeleton-texture separation will be derived. The additive separation-model means that the image $I(x,y)$ can be represented as the sum of the skeleton image $U(x,y)$ and the texture image $V(x,y)$ as follows:

$$I(x,y) = U(x,y) + V(x,y). \tag{4}$$
$$, U(x,y) = r(x,y) \cdot \left(\mathbf{N}_S(x,y)^T \cdot \mathbf{L}\right)$$
$$V(x,y) = r(x,y) \cdot \left(\mathbf{N}_S(x,y)^T \cdot \mathbf{P}^T(x,y) \cdot \mathbf{L}\right)$$

In the additive separation-model, there is a cross-correlation between the skeleton image $U(x,y)$ and the texture image $V(x,y)$, and the two factors U, V to be separated from I are not mutually independent. Hence, it is not easy to separate U and V accurately according to the additive separation-model. The existing methods for the skeleton-texture separation are based on the additive separation-model, and they do not necessarily achieve satisfactory separation for really observed images [10] in that the separated factors U, V are correlated to each other.

2.2.2 Multiplicative Model and Skeleton-Texture Separation

In the second case of the perturbed albedo, assuming that the albedo $r(x,y)$ within each object in the image is written as

$$r(x,y) = r_S \cdot \left(1 + s(x,y)\right), \tag{5}$$

where the function $s(x,y)$ is a perturbation function vibrating around zero and r_s is a given constant, then the multiplicative model for the skeleton-texture separation will be derived. The multiplicative separation-model means that image $I(x,y)$ can be represented as the product of the skeleton image $U(x,y)$ and the texture generator $V(x,y)$ as follows:

$$I(x,y) = U(x,y) \times V(x,y). \tag{6}$$
$$, U(x,y) = r_S \cdot \left(\mathbf{N}^T(x,y) \cdot \mathbf{L}\right), V(x,y) = 1 + s(x,y)$$

In this case, there is no cross-correlation between the skeleton image $U(x,y)$ and the texture generator $V(x,y)$, and the two factors U, V to be separated from I are independent of each other. Therefore, from the standpoint of the independence of the two factors U, V, the multiplicative separation-model is considered to be more rational than the additive separation-model.

To simplify the problem to separate the contributions of the skeleton image U from those of the texture generator V, we should convert the image brightness I to the logarithmic domain, so that the product of (6) is turned into a sum,

$$f = u + v, \tag{7}$$
$$f = \log I,\ u = \log U,\ v = \log V.$$

2.3 Skeleton-Texture Separation Algorithm

The separation problem based on the log-transformed model of (7) is formulated as the recovery problem that we separate the two contributions, the log skeleton-image u and the log texture-generator v, given the log image-brightness f. This separation problem is a typical ill-posed problem. The separation problem takes just the same form with the separation problem based on the additive separation model of (4). Vese and Osher [9] proposed an algorithm to solve the skeleton-texture separation problem based on the additive separation-model of (4). Applying the Vese-Osher algorithm to the separation problem of (7), we can separate the log skeleton-image u and the log texture-generator v from the log image-brightness f. In the following, the Vese-Osher algorithm is briefly described.

The log skeleton-image u is modeled as being composed of multiple regions with smoothly-varying brightness divided by discontinuous boundaries, and the space of the log skeleton-image u is mathematically formulated as the BV (Bounded-Variation Function Space) [12]. The energy of the log skeleton-image u is defined as the TV (Total Variation) norm $J(u)$, as follows:

$$J(u) = \int \|\nabla u\| \, dx\, dy. \tag{8}$$

On the other hand, the log texture-generator v is modeled as an oscillatory function vibrating around zero, and the space of the log texture-generator v is formulated as the G space for oscillating pattern, which was recently introduced by Meyer [13]. The G space is very close to the dual space of the BV space [10]. The G space is defined as the Banach space consisting of all functions that can be written as

$$v(x,y) = \partial_x g_1(x,y) + \partial_y g_2(x,y), \tag{9}$$
$$g_1, g_2 \in L^\infty\left(\mathbb{R}^2\right),$$

where the two functions g_1, g_2 are referred to as the oscillating mother functions. The energy of the log texture-generator v is defined as the G norm $\|v\|_G$, as follows:

$$\|v\|_G = \inf_{g_1, g_2} \left\{ \left\| \sqrt{(g_1)^2 + (g_2)^2} \right\|_{L^\infty} ; v = \partial_x g_1 + \partial_y g_2 \right\}, \tag{10}$$

where the infimum is computed over all possible decompositions (9) of v.

Under the above assumptions, the separation problem of (7) is formulated as the variational problem minimizing the following energy functional $E(u,v)$:

$$E(u,v) = J(u) + \mu \cdot \|v\|_G + \lambda \cdot \|f - u - v\|_2^2, \tag{11}$$

where λ, μ are positive parameters to be tuned optimally and the third energy term is a data-fidelity term that ensures f is very close to the sum of u and v. However, this separation algorithm is quite difficult to compute, because the computation of the G norm is not easy. To make the variational problem of (11) easier to solve, Vese and Osher replaced the G norm by the negative Sobolev norm, and reformulated the separation problem as the variational problem minimizing the following energy functional $F(u, g_1, g_2)$:

$$F(u, g_1, g_2) = \int \|\nabla u\| dx\, dy + \mu \cdot \int \sqrt{(g_1)^2 + (g_2)^2}\, dx\, dy$$
$$+ \lambda \cdot \int \left(f - u - \partial_x g_1 - \partial_y g_2\right)^2 dx\, dy. \tag{12}$$

The corresponding Euler-Lagrange equations are as follows:

$$u = f - \partial_x g_1 - \partial_y g_2 + \frac{1}{2\lambda} div\left(\frac{\nabla u}{|\nabla u|}\right), \tag{13}$$

$$\mu \frac{g_1}{\sqrt{g_1^2 + g_2^2}} = 2\lambda\left[\partial_x(u - f) + \partial_{xx} g_1 + \partial_{xy} g_2\right], \tag{14}$$

$$\mu \frac{g_2}{\sqrt{g_1^2 + g_2^2}} = 2\lambda\left[\partial_y(u - f) + \partial_{xy} g_1 + \partial_{yy} g_2\right]. \tag{15}$$

The skeleton image $U(x,y)$ is transformed from $u(x,y)$ as follows:

$$U(x,y) = \exp\left(u(x,y)\right), \tag{16}$$

whereas the texture generator $V(x,y)$ is transformed from $g_1(x,y)$, $g_2(x,y)$ as follows:

$$V(x,y) = \exp\left(v(x,y)\right) = \exp\left(\partial_x g_1(x,y) + \partial_y g_2(x,y)\right). \tag{17}$$

The two reconstructed functions $U(x,y)$, $V(x,y)$ are then multiplied together to obtain the reconstruction of the input image $I(x,y)$. However, the reconstructed image brightness is not necessarily equal to the original image brightness. Then, we define a residual image $D(x,y)$ as follows:

$$D(x,y) = I(x,y) - U(x,y) \cdot V(x,y). \quad (18)$$

The residual image D captures random textures and reconstruction errors.

As to the extension of the separation algorithm to color images, in this paper we employ the most straightforward approach in which all the color channels are treated independently.

2.4 Separation Results

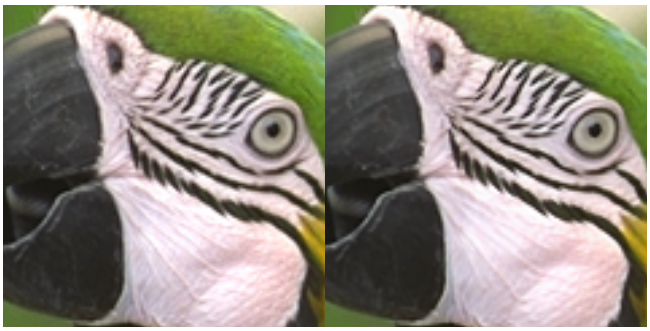
Separation simulations using natural color images show that the skeleton-texture separation method based on the multiplicative model is superior to that based on the additive model in that faster convergence is achieved and parameter setting is much easier. Fig.1 and Fig.2 show the separated skeletons U and the separated texture-generators V given by the skeleton-texture separation method based on the multiplicative model. In these figures, the recovered image $U \cdot V$ is also shown. In Fig.1 and Fig.2, textures are almost completely removed from the skeleton images in which objects' boundaries are sharp and not blurred at all.

3. IMAGE INTERPOLATION BASED ON THE SKELETON-TEXTURE SEPARATION

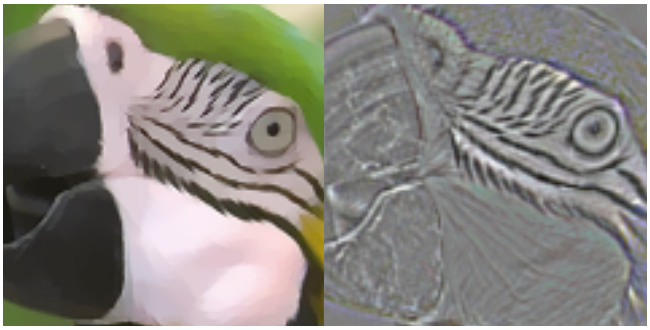
The basic idea of our image interpolation approach is as follows. Applying the skeleton-texture separation algorithm based on the multiplicative separation-model to an image I to be interpolated, first we decompose the image I as follows:

$$I(x,y) = U(x,y) \cdot V(x,y) + D(x,y). \quad (19)$$

Then we interpolate each of the three ingredients independently



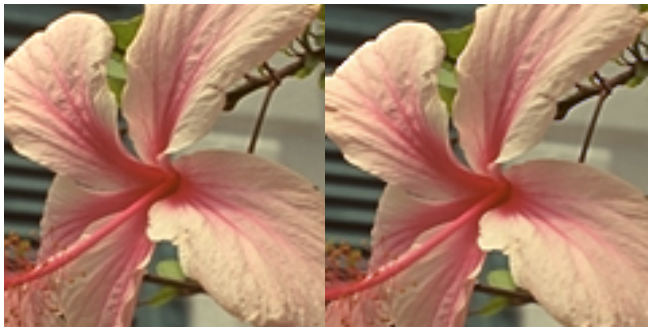
(a) Input image (b) Recovered image $U \cdot V$ $PSNR$ = 37.47 [dB]



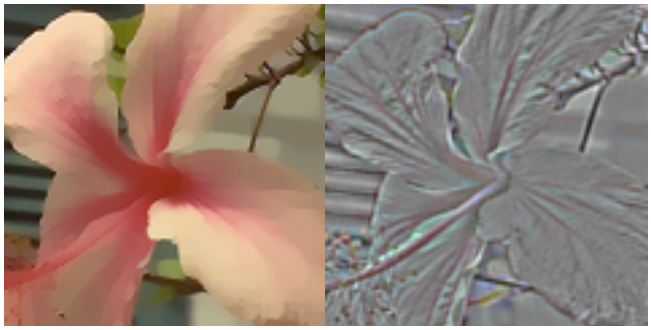
(c) Skeleton image U (d) Texture generator $200 \cdot (V\text{-}1.0)+128$

(d) Residual image D+128

Figure 1 - Skeleton image U, the texture generator V, the residual image D and the recovered image given by the method based on the multiplicative model (λ=2.0, μ=0.01)



(a) Input image (b) Recovered image $U+V$ $PSNR$ = 38.23 [dB]



(c) Skeleton image U (d) Texture generator $200 \cdot (V\text{-}1.0)+128$

(d) Residual image D+128

Figure 2 - Skeleton image U, the texture generator V, the residual image D and the recovered image given by the method based on the multiplicative model (λ=2.0, μ=0.01)

with a proper interpolation method suitable to each ingredient. Producing magnified ingredients, *U'*, *V'*, *D'*, we reconstruct an interpolated image *I'* as follows:

$$I'(x,y) = U'(x,y) \cdot V'(x,y) + D'(x,y). \quad (20)$$

In this reconstruction process, we can take account of use's taste in picture quality. When a user prefers a sharp image with high contrast, for instance first the texture generator *V* is enhanced and then an interpolated image *I'* is produced as follows:

$$I'(x,y) = U'(x,y) \cdot V_\alpha'(x,y) + D'(x,y) \quad , \alpha \geq 1. \quad (21)$$

$$V_\alpha' = \text{Interpolation}\left[V^\alpha\right] = \text{Interpolation}\left[\exp(\alpha \cdot v)\right]$$

On the other hand, when a user prefers a soft image with moderate contrast, for instance the interpolated residual image *D'* is reduced and thus an interpolated image *I'* is produced as follows:

$$I'(x,y) = U'(x,y) \cdot V'(x,y) + \beta \cdot D'(x,y) \quad , 0 \leq \beta \leq 1. \quad (22)$$

We use a proper interpolation method suitable to each gradient in the multiplicative separation-model. The texture generator *V* is an oscillatory function representing regular clear textures, and hence its proper interpolation method is a standard linear interpolation method. The residual image *D* is a function representing irregular faint textures and reconstruction errors, and hence its proper interpolation method is a statistical re-sampling interpolation method. The statistical re-sampling method first computes a mean and a variance of residual signals within a local neighboring region around a pixel to be interpolated, samples a Gaussian random variable with the mean and the variance and then inserts it into *D'* as an interpolation value of the pixel.

On the other hand, the skeleton image *U* corresponds to a cartoon approximation of the image *I* and lives in the bounded-variation function space. Hence, its proper interpolation method should preserve discontinuous jumps without producing blurs and ringing artifacts near the jumps. To suppress the ringing artifacts, one of the most effective approaches is to introduce into the interpolation approach the idea of the super-resolution restoring spatial frequency components higher than the Nyquist frequency. Recently, Malgouyres and Guichard [5] formulated the deblurring-oversampling problem with the super-resolution as a variational problem with the TV energy [12]. They introduced the sub-sampling into the TV variational method for the deblurring, and studied the minimization of the energy functional:

$$G(U') = \left\| \rho * U'(m,n) - U_{m,n} \right\|_2^2 + \gamma \cdot J(U') \quad , \gamma > 0. \quad (23)$$

They proved that for a cylindrical function whose DFT is supported by a line, the minimization of the energy functional of (23) admits a solution cylindrical along the same line. The cylindrical function mathematically models the 1-D structure of a step edge. The TV-based deblurring-oversampling method of (23) restores frequency components higher than the Nyquist frequency from observed blurry frequency components so that it can enlarge skeleton images while preserving structures defined as the cylindrical functions, and thus it can preserve discontinuous jumps without producing ringing artifacts. Therefore, this super-resolution deblurring-oversampling method is the most proper for the interpolation of skeleton images.

4. INTERPOLATION RESULTS

Fig.3 shows interpolated images given by our adaptable image interpolation method. We produce interpolation images, shown in Fig. 3, by changing the parameter α in the interpolation scheme of (21). In Fig.3, for comparison, images interpolated by the pixel replication method, the bicubic method and the super-resolution deblurring-oversampling method [5] are shown. As shown in Fig.3, our interpolation method produces sharper edges and suppresses jaggy artifacts better than the two classical methods, the pixel replication method and the bicubic method. When instead of separating skeleton and texture from an image we apply the super-resolution deblurring-oversampling method directly to the image interpolation, as shown in Fig.3(d) fine textures tend to be eliminated from its recovered image, and the interpolated image gives unnatural flat impression. On the other hand, our interpolation method preserves fine textures. Furthermore, controlling the parameter α, our interpolation method can adjust edge sharpness and texture contrast to reconstruct a high quality image according to user's taste in picture quality.

5. CONCLUSIONS

We make it possible to use interpolation algorithms suitable to the skeleton image, the texture generator, and the residual image respectively, and thus form a high-quality image interpolation method that controls edge sharpness and texture intensity according to user's taste in picture quality. Simulations conducted on real images show the merits of our interpolation approach.

REFERENCES

[1] M. Unser, "Splines: a perfect fit for signal and image processing," *IEEE Signal Processing Magazine*, 16, 11, pp.22-38, 1999.

[2] J. Allebach and P.W. Wong, "Edge-directed interpolation," in *Proc. IEEE Int. Conf. on Image Processing*, pp.707-710, 1996.

[3] X. Li and M.T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Processing*, 10, 10, pp.1521-1527, 2001.

[4] J.A. Leitao, M. Zhao and G. De Haan, "Content-adaptive video up-scaling for high-definition displays," in *Proc. SPIE Image and Video Communications and Processing*, 5022, pp.612-622, 2003.

[5] F. Malgouyres and G. Guichard, "Edge direction preserving image zooming: a mathematical and numerical analysis," *J. Num. Anal.*, 39, 1, pp.1-37, 2001.

[6] B.S. Morse and D. Schwartzwald, "Image magnification using level-set reconstruction," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp.333-340, 2001.

[7] H.Q. Luong and W. Philips, "Sharp image interpolation by mapping level curves," in *Proc. SPIE Visual Communications and Image Processing*, 5960, pp.2012-2022, 2005.

[8] K. Ratakonda and N. Ahuja, "POCS based adaptive image magnification," in *Proc. Int. Conf. Image Processing*, 3, pp.203-207, 1998.

[9] L. Vese and S. Osher, "Modeling textures with total variation minimization and oscillating patterns in image processing," *Journal of Scientific Computing,* 19, pp.553-572, 2003.

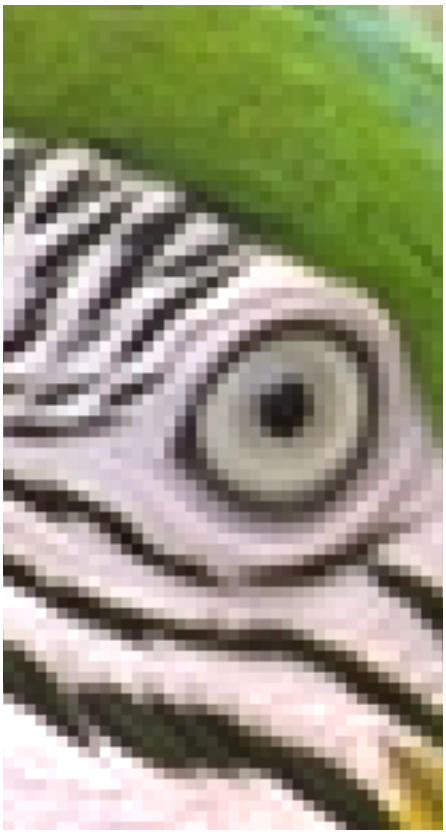
[10] J.F. Aujol and A. Chambolle, "Dual norms and image decomposition models," *Int. J. Computer Vision*, 63, 1, pp.85-104, 2005.

[11] P.T. Eliason, L.A. Soderblom and P.S. Chavez, "Extraction of topographic and spectral albedo information from multi spectral images," *Photogram Eng. Remote Sensing*, 48, pp.1571-1579, 1981.

[12] L. Rudin, S. Osher, and E. Fetami, "Nonlinear total variation based noise removal algorithm," *Physica D*, 60, pp.259-268, 1992.

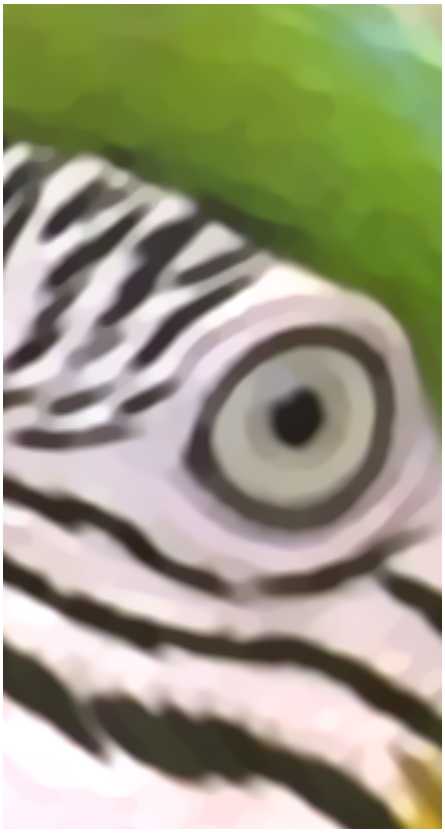
[13] Y. Meyer, *Oscillating patterns in image processing and nonlinear evolution equations: the fifteenth Dean Jacqueline B. Lewis memorial lectures*, University Lecture Series, 22, American Mathematical Society, 2001.

(a) Input image

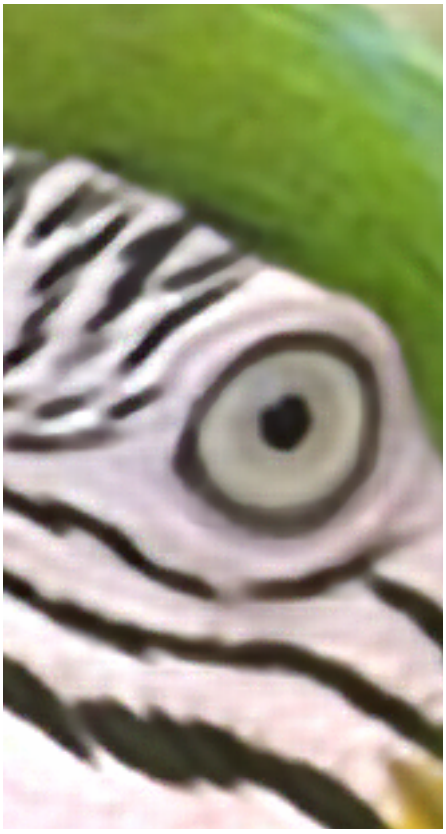


(b) Interpolated image by the pixel replication method

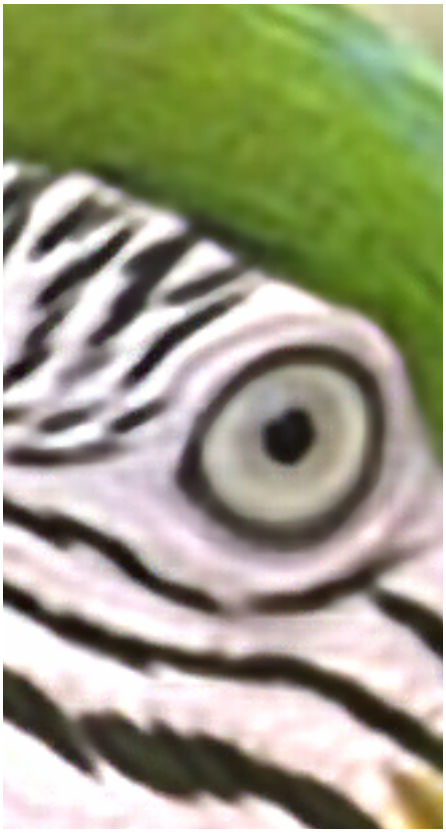
(c) Interpolated image by the bicubic method



(d) Interpolated image by the super-resolution deblurring-oversampling method [5]



(e) Interpolated image by our method of (21): α=1.0



(f) Interpolated image by our method of (21): α=1.4

Figure 3 - Portions of the parrot image magnified 4x using our interpolation method