SUBJECTIVE EVALUATION OF SINGLE-FRAME SUPERRESOLUTION ALGORITHMS[1]

Hüseyin Özdemir, Bülent Sankur

Electrical and Electronic Engineering Department, Boğaziçi University
huseyinozdemir@yahoo.com, bulent.sankur@boun.edu.tr

ABSTRACT

In this paper, we comparatively evaluate several single-frame interpolation methods with respect to a reference multi-frame superresolution algorithm. The objective performance was measured using Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM), while subjective evaluation was based on two-alternative forced choice scores. Rank correlation was used to assess the degree of agreement between subjective and objective measures, and to determine the best performing single-frame interpolation method to be used in video industry.

1. INTRODUCTION

Superresolution (SR) techniques aim to augment the resolution level of an image or a video with post-processing. While there is a plethora of SR algorithms, the video industry is still grappling with a viable method which is also feasible with the present technology.

There are two paradigms in superresolution: The first paradigm creates a high-resolution (HR) image using more than one low-resolution (LR) image. These multiframe SR image reconstruction methods fuse data coming from different LR images of the same scene, which however must differ at sub-pixel shifts. The second paradigm uses a single LR image to create a high-resolution (HR) version. These methods exploit various image models.

In [10], Ouwerkerk compared the subjective and objective performance of seven image interpolation methods. The characterizing property of these methods are as follows: interpolation kernel, sum of interpolation kernels with fuzzy weights; feedforward neural network using spatial error terms which consist of pixel averages, of edges and of line structures; local correlation-based interpolation; anisotropic diffusion; super resolution exploiting the sparse derivative prior; edge-directed interpolation; and finally, locally adaptive zooming algorithm. His objective tests based on PSNR, SSIM and edge stability metrics indicate that resolution synthesis algorithm performed best. Subjective tests, using edge blurring, edge blocking and generation of detail, both resolution synthesis and local correlation super resolution showed good results.

Video SR from a single frame is attractive due to its reduced complexity in video industry. We conjecture that single-frame interpolation methods have a niche role in video industry in that the viewer satisfaction would be on a par with respect to that obtained with more complicated spatio-temporal algorithms. With this purpose in mind, we evaluate both the objective and subjective performance of several single-frame interpolation methods vis-à-vis a well-known multiframe algorithm.

The single-frame interpolation methods list as follows: 1) Bicubic interpolation [5]; 2) Wavelet-based interpolation [6]; 3) Edge-adaptive interpolation [7]; 4) Markov-model interpolation [8]. These single-frame methods are compared with a multiframe method by Elad and Feuer [2]. The rationales for the choice of these methods were: i) Enable implementation at video rates; ii) Experiment with different methods. In fact, while the first three of the methods belong to the paradigms already used in Ouderwerk [10] (though the methods themselves are different), the last method is a new one.

The contribution of our paper vis-à-vis previous comparative survey papers is twofold: 1) We extend the comparisons to video data; 2) Subjective tests address issues frame-based versus video-based preferences in addition to correlations between objective and subjective scores. In Section 2 of the paper we present briefly the super-resolution (SR) algorithm, and in Section 3 the four image interpolation algorithms. Section 4 discusses the outcome of subjective and objective performance measurements, and conclusions are drawn in Section 5.

2. SR IMAGE RECONSTRUCTION USING LMS ALGORITHM

In [2], [3] and [4] the Least Mean Square (LMS) filtering method is developed for SR reconstruction. Elad and Feuer [2] formulated the super-resolution reconstruction algorithm for continuous image sequences. Let ideal image sequence be related to degraded image sequence via the relationship

$$Y(t-k) = DH(t-k)F(t,k)X(t) + E(t,k) \qquad (1)$$

where $Y(t-k)$ is degraded image vector at time $-\infty \le t-k \le t$ while $X(t)$ is the ideal SR image vector, all in lexicographical notation. D, $H(t-k)$ and $F(t,k)$ denote decimation operator, linear-space-variant blur operator, and backward geometric warp operator, respectively.

[1] This work was supported by Boğaziçi Univ. project 08A202 and TUBITAK project 107E001

We have the warp operator F using full-search block-matching algorithm (block size 16x16). Finally, $E(t,k)$ is the additive Gaussian noise vector with autocorrelation matrix $W^{-1}(t,k)$. Under the assumption that all these matrices are known [2] the LMS error is defined as

$$\varepsilon^2(t) = \sum_{k=0}^{\infty} \left\| Y(t-k) - DH(t-k) F(t,k) \hat{X}(t) \right\|^2_{W(t,k)} \quad (2)$$

Estimation of ideal image vector $\hat{X}(t)$ minimizing above error term can be found using iterative techniques. Using the steepest descent rule [1]:

$$\hat{X}(t) = F^{+}(t,1) \hat{X}(t-1) - \frac{\mu}{2} \frac{\partial \varepsilon^2(t)}{\partial \hat{X}(t)} \bigg|_{F^{+}(t,1)\hat{X}(t-1)} \quad (3)$$

where $\hat{X}$ is the estimated SR image and $F^{+}(t,1)$ is the forward geometric warp operator, that is, the pseudo-inverse of $F(t,1)$. The recursive implementation of the LMS SR method is convenient for real-time video applications since only the previous frame need to be stored.

3. IMAGE INTERPOLATION ALGORITHMS

3.1. Bicubic Interpolation

Given a sampled signal, its continuous counterpart can be approximated using some suitable interpolation kernel. 2D interpolation is usually accomplished by applying successively 1D kernel interpolation on horizontal and vertical directions. For uniformly spaced data, the continuous-domain signal $Y(u,v)$ can be written as

$$Y(u,v) = \sum_k \sum_l y(k,l) h\left(\frac{u-u_k}{\Delta u} \right) h\left(\frac{v-v_k}{\Delta v} \right) \quad (4)$$

In this expression, $(\Delta u, \Delta v)$ are sampling intervals, $h()$ is the interpolation kernel and {y(k,l)} represent the pixel array in the LR grid. The SR signal is obtained by resampling (4) on a finer grid. In [5], the cubic convolution kernel is given as

$$h(s) = \begin{cases} 1.5|s|^3 - 2.5|s|^2 + 1 & 0 \le |s| < 1 \\ -0.5|s|^3 + 2.5|s|^2 - 4|s| + 2 & 1 \le |s| < 2 \\ 0 & 2 \le |s| \end{cases} \quad (5)$$

3.2 Wavelet-Based Image Interpolation

Wavelet-based image interpolation methods assume that the available image is the coarse approximation (LL_0), that is, low-pass filtered subband of an HR image. The interpolation methods then first try to recover the missing horizontal (LH_0), vertical (HL_0) and diagonal (HH_0) detail subbands, and then obtain the HR image by taking the inverse Discrete Wavelet Transform (DWT) of the expanded image. An important property of DWT is the persistence property. In fact several wavelet-based compression schemes, such as embedded zero tree wavelets, employ this property. Temizel and Vlachos [6] proposed a wavelet based image interpolation method. They used the idea of "persistence", which implies that the magnitudes of wavelet coefficients corresponding to the same spatial location tend to propagate from lower scales to higher resolution scales. They extended the "persistence" idea to correlation coefficients. First, one goes one scale down, and estimates HL_1 by high pass filtering LL_1 horizontally. Then, correlation coefficients between HL_1 and its estimate are computed. Using these correlation coefficients and estimate of HL_0, exact value of HL_0 is computed. All these horizontal and vertical filtering operations are, however, implemented without decimation, in other words one stays at the resolution level of LL_0. The HH_0 is not predicted since it is judged visually less informative, so that the corresponding coefficients in the inverse DWT are filled with zeros.

3.3 Edge Adaptive Image Interpolation

The imaging process and the concomitant loss of resolution are modeled as low-pass filtering and decimation stages in [7]. The low-pass filtering operation modifies the values of the pixels near the edges proportionally to the distance between pixels and the edge. Therefore the analysis of the low resolution pixels should give an idea about the position of the edge at sub-pixel level. The one-dimensional case is illustrated in Fig. 1.

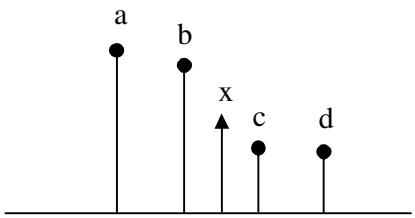

Figure 1 – The x pixel, astride an edge, is to be interpolated.

The interpolated value x between the given a, b, c, d neighbors become [7]:

$$x = \mu b + (1-\mu) c \qquad \mu = \frac{k(c-d)^2 + 1}{k\left((a-b)^2 + (c-d)^2\right) + 2} \quad (6)$$

where k is an input parameter and affects the edge sensitivity. When edge is in midway between b and c, $a-b=c-d$ and $x=(b+c)/2$. When edge is closer to c, then $a-b<c-d$ and x takes a value closer to b. In 2D, first the missing pixels along the rows and the missing pixels along the columns are estimated, separately. Then the diagonal pixels are estimated using the results of the previous steps and the mean value of the two results is taken.

3.4 Deterministic-Stochastic Interpolation

Nemirovsky and Porat [8] have proposed an image interpolation method where texture is assumed to be composed of two orthogonal components: a purely non-deterministic component and a deterministic component. The deterministic part of the HR image is computed by calculating the periodogram of the LR image, finding the peaks of the periodogram, masking out the rest and using inverse DFT with zero padding. The non-deterministic part starts with the LR image whose deterministic part has

been removed. After zero padding the LR non-deterministic part of the image, the missing pixels are filled as described below. For a discrete random field *Y*, the linear least square estimate of *Y(m,n)* can be found as the linear combination of the values in the causal neighborhood of *(m,n)*.

$$\hat{y}(m,n) = \sum_{\substack{i,j \\ (m-i,n-j)\in D_{m,n}}} c_{i,j}\, y(m-i,n-j) \qquad (7)$$

where the coefficients are found via minimum mean square approach. In [8], 3 nearest neighbors are considered, namely $D = \{(0,1),(1,1),(1,0)\}$. The HR image is simply the sum of the HR deterministic and non-deterministic parts.

4. EXPERIMENTAL RESULTS

The original RGB video material had 704x480 pixel resolution; they were decimated by factor of 2 using bilinear interpolation to CIF resolution and converted to grayscale *Y* using the equation below $Y = 0.299R + 0.587G + 0.114B$. Then, the decimated sequences were resized to their original resolution (704x480) by using the methods described above. We used 3 video sequences (Calendar, Susi, Tennis), each containing 250 frames. Error images for the 100th frames of Susi sequence are given in Fig. 2. The error images were found by subtracting the processed image from the original image, taking the absolute value of the result and scaling by 10. We also extracted two frames from each video for still frame evaluation. PSNR and SSIM metrics were used to evaluate the results quantitatively [9]. PSNR was calculated directly for still frames. For video PSNR calculation, first mean square error (MSE) of each frame in the video was calculated and these MSE values were averaged to find the average MSE value of the video. Finally average MSE value was used to find the video PSNR. Video SSIM values were calculated as described in [9]. Measurements are given in Table 1. According to PSNR criteria, LMS and bicubic methods have the same scores in video and still frame cases. According to SSIM, the winner is LMS method, which, performs better than single-frame methods.

4.1 Subjective Preferences

Subjectively perceived quality of the videos and single frames produced by various superresolution methods were evaluated based on the two-alternative forced choice (2AFC) paradigm. A total of 28 people took part in the experiment. Subjects had normal or corrected-to-normal vision and their ages varied between 22 and 30. The experiments were performed in a lit room using a 60 Hz, 1280x720 LCD display from a viewing distance of 30 cm. Each observer completed a total of 10 2AFC comparisons for each of the three sequences, 10 2AFC comparisons for each of the six frames, in total 90 2AFC comparisons during the experiment.

	Video PSNR	Still Frame PSNR	Video SSIM	Still Frame SSIM
LMS	24.9	24.9	0.85	0.86
Bicubic	24.9	24.9	0.85	0.85
Edge Adap.	23.7	23.8	0.81	0.82
Wavelet	22.9	23.0	0.77	0.78
MRF Pre.	24.4	24.4	0.77	0.77

Table 1 – Average PSNR (dB) and SSIM values.

The ten comparisons are due to the two-combinations of the 5 methods. There was no time pressure on the subjects and they could toggle back and forth between test pairs as much as they wanted. Subject votes were screened for outlier behavior and only one was eliminated.

4.2 Correlation of video and frame scores

The percentages of subjective preferences are shown in Table 2. In each cell, the upper number is the video score and lower one is the frame score. For example, the second cell in the first row of Table 2 indicates that SR method is preferred over the bicubic 75.3% (in video tests) of the time, while the third cell indicates that it is overwhelmingly preferred over the edge-adaptive method, and so on. Note that the scores are normalized, so maximum is 100 for each comparison. We wanted to check if the subjective preferences for the methods were in agreement between the frame tests and video tests, and this agreement was measured with Spearman's rank correlation coefficient. If the number of comparisons is K (which is given by combinatorial two of the number of methods), each subject form a K dimensional vector for each frame and video. Each element of this K dimensional vector is related to "method A vs. method B" comparison. Combining the decisions of all subjects for this comparison, we get 3 vectors in sequence case and 6 vectors in frame case. Then, 3 sequence vectors are averaged and the resultant vector represents the decisions related to "method A vs. method B" comparison in sequence case. Also, 6 sequence vectors are averaged and the resultant vector represents the decisions related to "method A vs. method B" comparison in still frame case. The average sequence and the average still frame vectors are used to compute Spearman's rho coefficients which are given in Table 3. The correlation coefficients, except for those of edge adaptive – MRF prediction and bicubic – wavelet pairs, are high. This proves that in most cases frames or sequences can be used alternatively for judging superresolution methods.

4.3 Correlation of objective and subjective preferences

The ranking of methods according to both subjective and objective preferences (PREF) is given in Table 4.

100th frame of 1st video

Bicubic interpolation

Edge adaptive

Interpolation using MRF

LMS

Wavelet based interpolation

Figure 2 – Error images of all methods for 100th frame of 1st video.

	LMS	Bicubic	Edge Adap.	Wavelet	MRF Pre.
LMS		75.3 97.5	96.3 96.3	87.7 90.1	98.8 98.8
Bicubic	24.7 2.5		92.6 85.2	67.9 46.3	92.6 88.9
Edge Adap.	3.7 3.7	7.4 14.8		11.1 19.1	58.0 56.2
Wavelet	12.3 9.9	32.1 53.7	88.9 80.9		91.4 85.8
MRF Pre.	1.2 1.2	7.4 11.1	42.0 43.8	8.6 14.2	

Table 2 – Subjective preferences in percentage for both videos and frames.

	LMS	Bicubic	Edge Adap.	Wavelet	MRF Pre.
LMS		1.0	1.0	1.0	1.0
Bicubic	1.0		0.99	0.54	1.0
Edge Adap.	1.0	0.99		0.66	0.48
Wavelet	1.0	0.54	0.66		0.62
MRF Pre.	1.0	1.0	0.48	0.62	

Table 3 – Spearman's rank correlation between subjective video and frame preferences.

	Video PSNR	Frame PSNR	Video SSIM	Frame SSIM	Video Subj.	Frame Subj.
LMS	2	2	1	1	1	1
Bicubic	1	1	2	2	2	3
Edge Adap.	4	4	3	3	4	4
Wavelet	5	5	5	4	3	2
MRF Pre.	3	3	4	5	5	5

Table 4 – Rankings of methods according to subjective and objective scores.

Rankings for PSNR and SSIM measures were obtained simply by converting these figures (Table 1) to ranks. Rankings for PREF scores were obtained by summing the scores along rows of Table 2 and converting the resulting totals to ranks. Note the high degree of parallelism between these rank results. Spearman's rank correlation coefficients for PREF-vs-PSNR and PREF-vs-SSIM scores were computed. To this effect three 90-element vectors were created for PSNR, SSIM and PREF measurements, since there are over all 90 instances of comparisons. For example, the first element of this vector can correspond to the comparison of LMS-vs-Bicubic method as tested on Susi video; and so on for other vector elements. Each element results from the rounded average of all 28 subjects. Similarly, 90-long SSIM and PSNR vectors were created. If method A had a higher PSNR (SSIM) value as compared that of method B, the corresponding cell was assigned the value 1, and otherwise 0. These three vectors were used to compute Spearman's rank correlation coefficients between subjective and objective measurements. These rank correlations were found as 0.8676 between PREF and PSNR and as 0.8875 between PREF and SSIM measurements. Method to method comparisons are given in Table 5. In each cell, the upper number is PSNR related coefficient and lower one is SSIM related coefficient. The correlation between subjective preferences and both SSIM and PSNR measurements is quite high. This implies that tedious subjective tests can be replaced by objective SSIM measurements. Notice that the smallest correlation occurs in the LMS – Bicubic cell of Table 5. The result of bicubic method resembles the original input. So, bicubic method gets high objective measurements (PSNR and SSIM results). On the other hand, the result of the LMS method is better than the original inputs, because of the fact that LMS is a multi-frame method and LR samples coming from neighboring frames increases the quality of the SR estimate. For this reason, LMS gets high subjective scores.

5. CONCLUSIONS

Experiments have indicated that the best-performing single-frame SR method was the bicubic method. This is also a good candidate to be employed in video industry given its implementation simplicity.

	LMS	Bicubic	Edge Adap.	Wavelet	MRF Pre.
LMS		0.30 -0.15	0.75 1.0	1.0 1.0	0.67 1.0
Bicubic	0.30 -0.15		1.0 1.0	1.0 1.0	1.0 1.0
Edge Adap.	0.75 1.0	1.0 1.0		1.0 1.0	0.9 1.0
Wavelet	1.0 1.0	1.0 1.0	1.0 1.0		0.87 0.87
MRF Pre.	0.67 1.0	1.0 1.0	0.9 1.0	0.87 0.87	

Table 5 – Correlation coefficients between PREF, SSIM and PSNR scores.

REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall Inc., 1986.

[2] M. Elad and A. Feuer, "Super-resolution restoration of continuous image sequence using the LMS algorithm," in *Eighteenth Convention of Electrical and Electronics Engineers*, Israel, March 7-8. 1995, pp. 2.2.5/1 - 2.2.5/5.

[3] G.H. Costa and J.C.M. Bermudez, "Statistical Analysis of the LMS Algorithm Applied to Super-Resolution Image Reconstruction," *IEEE Transactions on Signal Processing*, vol. 55, Issue 5, Part 2, pp. 2084 – 2095, May 2007.

[4] G.H. Costa and J.C.M. Bermudez, "Informed Choice of the LMS Parameters in Super-Resolution Video Reconstruction Applications," *IEEE Transactions on Signal Processing*, vol. 56, Issue 2, pp. 555 – 564, Feb. 2008.

[5] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, pp. 1153 – 1160, Dec 1981.

[6] A. Temizel and T. Vlachos, "Wavelet domain image resolution enhancement" *IEE Proceedings on Vision, Image and Signal Processing,* vol. 153, Issue 1, pp. 25 – 30, Feb. 2006.

[7] S. Carrato, G. Ramponi, S. Marsi, "A simple edge-sensitive image interpolation filter," *Image Processing, IEEE Proceedings,* International Conference, vol. 3, pp. 711 – 714, 16-19 Sept. 1996.

[8] S. Nemirovsky and M. Porat, "On Texture and Image Interpolation using Markov Models", *Signal Processing and Image Communication*, Elsevier, November 2008.

[9] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: From error measurement to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, Jan.2004.

[10] J.D. Ouwerkerk, "Image super-resolution survey", *Image and Vision Computing*, 24, pp. 1039-1052, 2006.