An Efficient Upsampling Technique for Images and Videos

Ming-Sui Lee and Chen-Wei Chang

Department of Computer Science and Information Engineering,
National Taiwan University, Taipei, Taiwan
mslee@csie.ntu.edu.tw, r96922080@ntu.edu.tw

Abstract. A block-based upsampling method for images and videos is proposed in this work. Block classification is first conducted in the DCT domain to categorize 8x8 image blocks into several types: smooth areas, edges and others. For the plain background and smooth surfaces, simple patches are used to enlarge the image size without degrading the resultant visual quality. Since human eyes are more sensitive to edges, a more sophisticated technique is applied to edge blocks. They are approximated by a facet model so that the image data at subpixel positions can be generated accordingly. By taking temporal information into account, this concept can further be applied to videos. To upsample an image block in the current frame, we may borrow the upsampled version of the corresponding block in the reference frame if the residual is tolerable. Experimental results are shown to demonstrate the great reduction of computational complexity while the output visual quality still remains satisfactory.

Keywords: Image/video upsampling, block-based processing.

1 Introduction

Since there are various multimedia terminals such as digital cameras, cellular phones, PDA's (personal digital assistants), computers and HDTV's available these days, there is a great need in developing techniques to facilitate flexible image/video format conversion. Quality degradation due to down-sampling, up-sampling, and coding/decoding in the transmission process is inevitable. Image interpolation or upsampling techniques together with enhancement have been studied for years.

Linear interpolation methods such as bilinear and bicubic interpolations are commonly used to enlarge an image due to their simplicity. However, they often result in blurred edges in the output image. Several complicated algorithms have been proposed to solve this problem. Set theoretic method with priori information, Projection onto convex sets (POCS), was proposed in 1989 by Stark et al [1]. Another approach of SR is Iterative Back-Projection (IBP) which was first introduced by M. Irani and S. Peleg in 1991[2]. A stochastic approach was proposed by Schultz and Stevenson in 1994[3]. That was a Bayesian framework which adopts the maximum a posteriori probability (MAP) to synthesize high resolution (HR) outputs. Overall speaking, those algorithms provide good results. However, they all suffer from extensive computation which is impossible to be utilized in real-time applications. Thus, the goal of

P. Muneesawang et al. (Eds.): PCM 2009, LNCS 5879, pp. 77–87, 2009.

this work is to propose an upsampling algorithm that dramatically reduces the processing time with acceptable visual quality.

We propose a content adaptive technique to upsample an image/video to an output image/video of higher resolution efficiently. It is a block-based processing algorithm. The most suitable up-sampling method is adaptively chosen for a block according to its type. This approach should be compared with the conventional method where a fixed processing technique is applied to the whole image. In the proposed algorithm, block classification is first conducted in the DCT domain to categorize each image block into several types: smooth areas, edges and others. For the plain background and smooth surfaces, simple patches are used to enlarge the image size without degrading the resultant visual quality. Since human eyes are more sensitive to edges, we adopt a more sophisticated technique to process edge blocks. That is, they are approximated by a facet model so that the image data at subpixel positions can be generated accordingly. For video upsampling, similar concepts are adopted to I-pictures. Along with motion vectors and residuals, we can further reduce the complexity while dealing with P- and B-pictures. The efficiency of the proposed technique is demonstrated by experimental results.

This paper is organized as follows. Block classification based on the DCT coefficients is investigated in Sec. 2. The content-adaptive image up-sampling algorithm is presented in Sec. 3. Experiments on video sequences are discussed in Sec. 4. Experimental results and discussion are shown in Sec. 5 followed by conclusion in Sec. 6.

2 Block Classification in the DCT Domain

An image may have heterogeneous contents like smooth areas, edges and textures. However, if an image is divided into blocks of a fixed size, the content in a block is more homogeneous if the block size is relatively not too big. Based on this observation, a content-adaptive image up-sampling method can be applied if image blocks are classified into different types appropriately. Block classification can be conducted in either the image pixel domain or the DCT coefficient domain. DCT-based block classification is needed when we process images encoded by the JPEG image compression standard. In this section, an efficient way to classify a block of size 8x8 to different types in the DCT domain is described. The two-dimensional (2D) DCT transform for an 8x8 block can be written as

$$F_{uv} = \frac{C_u C_v}{4} \sum_{i=0}^{7} \sum_{j=0}^{7} \cos\frac{(2i+1)u\pi}{16} \cos\frac{(2j+1)v\pi}{16} f(i,j) \tag{1}$$

where $C_k = \begin{cases} \frac{1}{\sqrt{2}} & if \quad k=0 \\ 1 & if \quad k>0 \end{cases}$ and $f(i,j)$ represents the image intensity value.

Thus, each DCT coefficient is a linear combination of 64 basis functions. Each basis function consists of different vertical and horizontal space frequencies. In other words, each DCT coefficient represents the energy of a specific pattern within that block. Therefore, it is possible to determine the block type based on the relative magnitudes of these DCT coefficients as detailed below.

The DC value represents the smoothness of a block while the first two AC values in the first column show the horizontal edges of lower spatial frequencies. Similarly, the first two AC coefficients in the first row represent the vertical edges and first two AC values in the diagonal represent the diagonal components. If most energy concentrates on the DC value, it means that this block is very likely a smooth surface. Furthermore, for those blocks that are not grouped as smooth area, we define a ratio that is the sum of AC values to the DC value. If the ratio is larger than a threshold, it means the corresponding 8x8 image block is dominated by edges. For those blocks which are classified neither a smooth area nor an edge block, they are categorized as others. Fig. 1 shows an example of block classification. As we can see from Fig. 1, a given image can be separated into three different groups successfully.

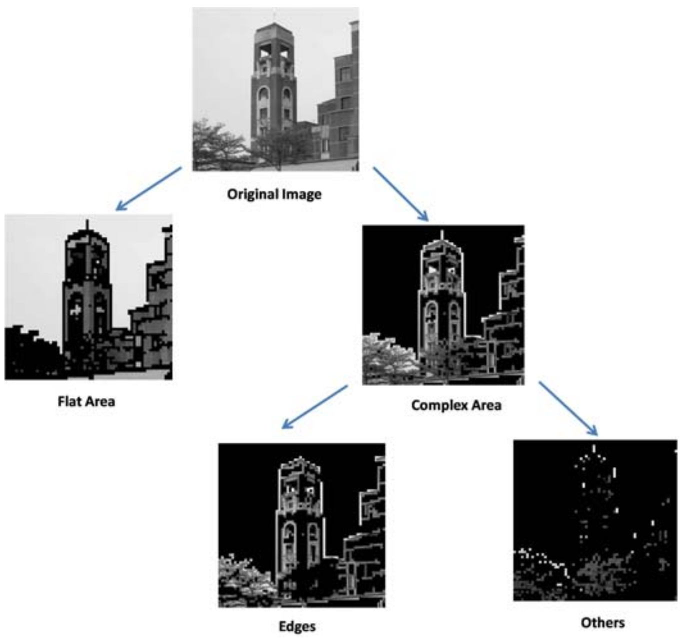


Fig. 1. An example of block classification of an image, "tower", with size 512x512. Blocks with original intensity represent that they are classified into different block types: flat areas, edges and others, respectively

3 Content-Adaptive Image Upsampling

Once an image is divided into different block types, content-adaptive upsampling techniques can be performed accordingly which are detailed as follows.

3.1 Upsampling of Edge Blocks

Since human eyes are more sensitive to edges, upsampling of edge blocks requires special treatment to guarantee the output visual quality. By viewing an image as a gray-level intensity surface, it can be approximated by a facet model. A facet model is built to minimize the difference between an intensity surface and observed image data. The piecewise quadratic polynomial is used in Haralick's facet model. That is, an image $F(j,k)$ is approximated by

$$\hat{F}(r,c) = k_1 + k_2 r + k_3 c + k_4 r^2 + k_5 rc + k_6 c^2 + k_7 rc^2 + k_8 r^2 c + k_9 r^2 c^2 \qquad (2)$$

where k_i are weighting coefficients to be determined and r and c are the row and column Cartesian indices of image $F(j,k)$ within a specified region. Least square solution is demanded to determine the coefficients, k_i, $1 \le i \le 9$. Noted that the solution of coefficients, k_i, is an ill-conditioned problem since polynomials $r^m c^n$, $m,n = 0,1,2$, are not orthogonal. To convert the ill-conditioned problem to a well-conditioned one, a set of orthogonal polynomials is adopted in the polynomial expansion instead. For example, we may consider the 3x3 Chebyshev orthogonal polynomials as given below:

$$\begin{aligned}
&P_1(r,c) = 1, \quad P_2(r,c) = r, \quad P_3(r,c) = c, \\
&P_4(r,c) = r^2 - \frac{2}{3}, \quad P_5(r,c) = rc, \quad P_6(r,c) = c^2 - \frac{2}{3} \\
&P_7(r,c) = c\left(r^2 - \frac{2}{3}\right), \quad P_8(r,c) = r\left(c^2 - \frac{2}{3}\right) \\
&P_9(r,c) = \left(r^2 - \frac{2}{3}\right)\left(c^2 - \frac{2}{3}\right)
\end{aligned} \qquad (3)$$

where $r,c \in \{-1,0,1\}$. As a result, the approximation can be rewritten in the form of

$$\hat{F}(r,c) = \sum_{n=1}^{N} a_n P_n(r,c) \qquad (4)$$

where a_n are polynomial coefficients which can be determined by the convolution of an image and a set of impulse response arrays. To obtain the facet model, observation equations are set up at integer parameters r and c to approximate the image value at a local region. When the facet model is applied to image upsampling, we compute $\hat{F}(r,c)$ at non-integer r and c values. It can be adopted to interpolate an image with any upsampling factor. For example, $\hat{F}(0.5,0.5)$ can be computed and inserted between $\hat{F}(0,0)$ and $\hat{F}(1,1)$ as shown in Fig. 2 so that the image can be enlarged by a factor of two. Similarly, the image size can be adjusted to any desired size by assigning different non-integer parameters such as $(0.25,0.25)$, $(0.3,0.3)$ into the approximating polynomial.

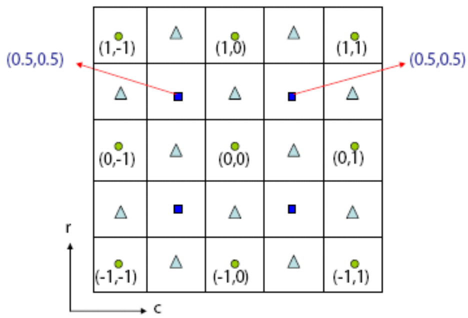


Fig. 2. The coordinates of a facet model

3.2 Upsampling of Non-edge Blocks

Blocks that belong to the plain background group contain smooth surfaces. Since there is not much variation in those areas, a bilinear interpolation method can be adopted to expand the image content without degrading the visual quality much. Bilinear interpolation followed by a technique called "unsharp masking" is applied to the rest of the blocks to enlarge the block size while magnifying the variations at the same time. This cascaded operation yields an output image block of good quality. The parameters of the unsharp mask, *e.g.* the size of the impulse response array and the weighting coefficients, control the sharpness of the output image. They can be chosen adaptively for different applications.

4 Content-Adaptive Video Upsampling

Consider MPEG sequences, each group of pictures (GOP) contains several picture types: I-picture, P-picture and B-picture. I-picture, serving as the basis of prediction for the whole GOP, is coded in DCT domain. The coding scheme for I-picture is similar to the JPEG image compression. Thus, the content-adaptive image upsampling technique described in the previous section can be performed to the luminance channel of I-pictures in a video sequence. Since human eyes are less sensitive to chrominance components, we just apply bilinear interpolation to chrominance channels in order to save computational complexity.

For P- and B-pictures, they are coded with motion vectors which are the displacements of blocks between the coded frames and reference frames. A block with a motion vector means that there exists a block in the reference frame that has similar content. In this case, rather than coding the whole image block, only residuals, the differences between current block in coded frame and reference blocks in reference frames, are stored. According to this temporal dependency of frames in a GOP, the upsampling algorithm can be applied even more efficiently.

If a block in the current coded frame is completely equal to another block in the reference frame, there is no need to apply any time-consuming SR algorithm to it again. In other words, if we have a block to be enlarged and this block has a corresponding block in the reference frame with same content, the upsampled version of

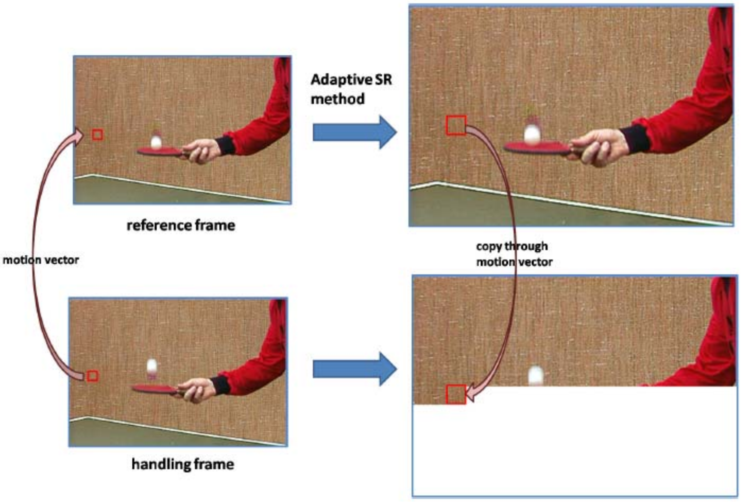


Fig. 3. An example that describes the concept of adopting temporal information to upsample a block with zero residual

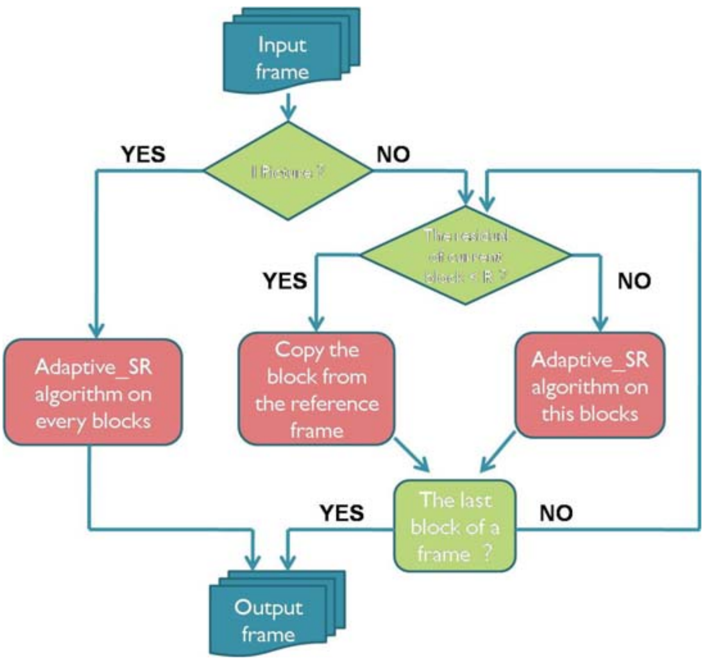


Fig. 4. The flow chart of the proposed upsampling algorithm to video sequences

the block in the reference frame can be extracted out directly to serve as the result of the block in the coded frame. Taking Fig. 3 for example, these two consecutive frames are very similar to each other. Almost all of the background areas in these two frames are the same; hence, when we enlarge the current handling frame, the blocks from the reference frame which are identical to the blocks in the handling frame are duplicated. As the result, the processing time can be reduced.

However, for a video sequence, the frame rate is 20~30 frames per second. Each frame is displayed for a short time, so we can loosen the criterion of the concept. That is, the blocks can be copied from the reference frame as long as the residual is lower than a threshold instead of setting residual strictly to be zero. For those blocks with large residual, the adaptive upsampling method is conducted according to the block type since copying the blocks in the reference frame degrades the visual quality. As a result, the computational complexity can be reduced greatly while the output visual quality still remains satisfactory. Fig. 4 shows the flow chart of the proposed upsampling algorithm to video sequences. Similar process is performed to all the blocks in a frame and all the frames in a sequence. There is a trade-off between the video quality and the processing time. Since the proposed block-based algorithm provides flexibility in choosing different upsampling methods and takes temporal dependency into account, users can adjust the parameters to fit their need adaptively.

5 Experimental Results

5.1 Single Image Upsampling

The experiments are performed on Intel(R) Core(TM)2 Duo CPU E6550 @ 2.33GHz, 2.00 GB RAM. In order to compare the quality and speed of each method, we display the image results and processing time of bilinear interpolation, MAP, Block-based_1 and Block_based_2, where Block_based_1 is the proposed content adaptive block-based method with facet modeling for edge blocks, and Block_based_2 is the proposed content adaptive block-based method with MAP for edge blocks. The resolution of original test images is 512x512. We first downsample original images by a factor of two both in vertical and horizontal directions, and then apply each upsampling algorithm to them. Table 1 and Table 2 show the processing time of different algorithms. The processing time of each method is also normalized with respect to bilinear interpolation. As we can see from these two tables, choosing different upsampling methods to different block types saves lots of computational complexity.

Table 1. Comparison of processing time for test image, "boat"

	Bilinear	**MAP**	**Block-based_1**	**Block-based_ 2**
Time (Sec)	0.046	0.764	0.078	0.250
Normalized	1	16.608	1.696	5.435

Table 2. Comparison of processing time for test image, "tower"

	Bilinear	**MAP**	**Block-based_1**	**Block-based_ 2**
Time (Sec)	0.046	0.750	0.061	0.202
Normalized	1	16.304	1.326	4.391

Fig. 5. Original image of size 512x512

Even though the processing time of block-based_2 (MAP for edge blocks) increases while comparing to block-based_1 (facet model for edge blocks), it still saves 60~75% computational cost according to image content complexity.

The visual quality comparison is shown in Fig. 5. Fig. 5 is the original image of size 512 x 512. The images upsampled by bilinear interpolation, MAP, block-based_1 and block-based_2 are demonstrated in Figs. 6(a), (b), (c) and (d), respectively. Clearly we can see from those images that bilinear interpolation provides poor visual quality and MAP and block-based_2 have similar performance. The proposed algorithm with facet modeling outperforms bilinear interpolation and provides comparable visual quality to MAP and block-based_2.

5.2 Video Upsampling

In this section, the proposed content-adaptive upsampling algorithm is performed to several test videos: ”foreman”, “akiyo”, “mobile”, “mother_daughter” and “pamphlet”. The original resolution of these videos is 352x288 with 300 frames in total and the GOP size is set to be 15 frames. Fig. 7(a) and Fig. 7(b) show the total processing time and the number of blocks that are copied from reference frame with various thresholds of residual. The higher the tolerable level of the residual, the more the blocks duplicated from reference frame. In other words, computational complexity can be reduced so that the processing time drops greatly. Note that the contents of two videos, “foreman” and “mobile”, have higher mobility introduced by camera motions so that they have relatively fewer blocks copied from the reference frame when compared with other three videos.

To further show the advantage of the proposed algorithm, the processing time is compared with two well-known methods, bilinear interpolation and MAP. As shown in Fig. 8, content-adaptive block-based method reduces the processing time successfully especially when the tolerable level gets higher and higher. Moreover, if compared with MAP, we can clearly see the big gap between these two methods.

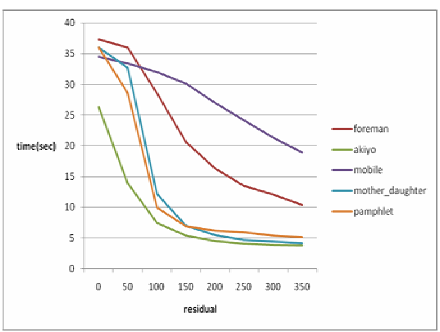
(a)

(b)

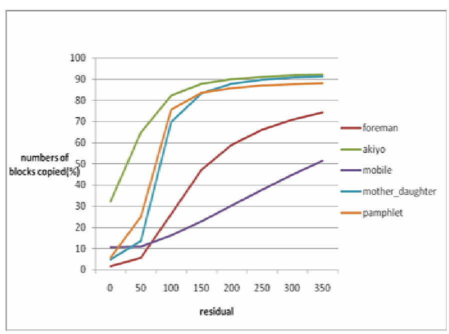
(c)

(d)

Fig. 6. Comparison of visual quality: (a) bilinear interpolation (b) MAP (c) block-based_1 and (d) block-based_2

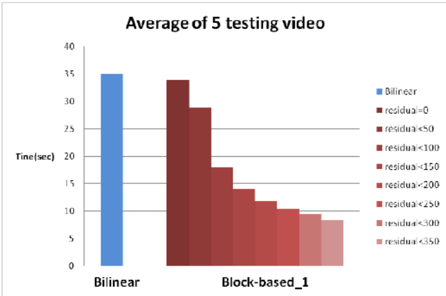


(a)

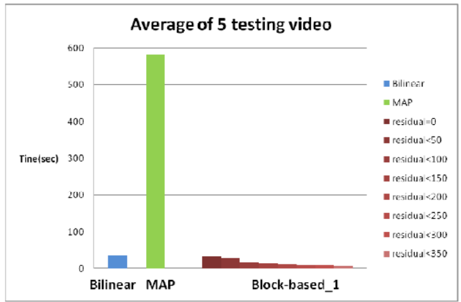


(b)

Fig. 7. Experimental results of five test videos. (a) the comparison of processing time and (b) the number of blocks copied from reference frame.



(a)



(b)

Fig. 8. The averaged processing time of five test videos. (a) compare with bilinear interpolation and (b) compare with MAP.

6 Conclusion

In this work, an image is divided into several block types according to the geometric property inherently in DCT coefficients. The proposed upsampling algorithm based on block classification is content-adaptive which adopts relatively low cost processing for blocks that contain less important information to save computational complexity for critical areas that require more sophisticated processing. Taking temporal dependency of video sequences into account, the proposed algorithm can also be performed efficiently to enhance the video resolution. It was shown by experimental results that the visual quality has been improved with sharper edges and the computational complexity has been reduced greatly while remaining satisfactory visual quality.

References

1. Stark, H., Oskoui, P.: High-resolution image recovery from image-plane arrays, using convex projections. J. Opt. Soc. Am. A 6, 1715–1726 (1989)
2. Irani, M., Peleg, S.: Improving resolution by image registration. Graphical Models and Image Processing 53, 231–239 (1991)
3. Schultz, R.R., Stevenson, R.L.: A Bayesian Approach to Image Expansion for Improved Definition. IEEE Transaction of Image Processing 3(3) (1994)
4. Gunturk, B.K., Altunbasak, Y., Mersereau, R.: Bayesian resolution-enhancement framework for transform-coded video. In: Proceedings of IEEE International Conference on Image Processing, vol. 2, pp. 41–44 (2001)
5. Gevrekci, M., Gunturk, B.K., Altunbasak, Y.: POCS-Based Restoration Of Bayer-Sampled Image Sequences. In: International Conference on Acoustics, Speech and Signal Processing (ICASSP), vol. 1, pp. I-753– I-756 (2007)
6. Tom, B.C., Katsaggelos, A.K.: Reconstruction of a High Resolution Image from Multiple Degraded Mis-Registered Low Resolution Images. In: SPIE VCIP, vol. 2308, September 1994, pp. 971–981 (1994)
7. Elad, M., Feuer, A.: Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. IEEE Trans. Image Processing 6(12), 1646–1658 (1997)

8. Liu, S., Bovik, A.C.: Efficient DCT-Domain Blind Measurement and Reduction of Blocking Artifacts. IEEE Trans. on Circuits and Systems for Video Technology 12(12), 1139–1149 (2002)
9. Pratt, W.K.: Digital Image Processing, 4th edn
10. Lee, M.-S., Shen, M.-Y., Jay Kuo, C.-C.: A content-adaptive up-sampling technique for image resolution enhancement. Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP) 1, 87–90 (2007)
11. Park, S.C., Park, M.K., Kang, M.G.: Super-resolution image reconstruction a technical overview. IEEE Signal Processing Magazine 20(3), 21–36 (2003)
12. Borman, S., Stevenson, R.L.: Super Resolution From Image Sequences - A Review. In: Proceedings of Circuit and Systems (1998)
13. Tudor, P.N.: MPEG-2 Video Compression. Electronics & Communication Engineering Journal (1995), http://www.bbc.co.uk/rd/pubs/papers/paper_14/paper_14.shtml
14. Isnardi, M.: MPEG-2 Video Compression, November 29 (1999), http://www.leitch.com/resources/tutorials/mpeg-2VideoCompression.pdf
15. Haskell, B.G., Puri, A., Netravali, A.N.: Digital Video: An Introduction to MPEG-2