A HVS-Guided Approach for Real-time Image Interpolation

Rui Chen, Huizhu Jia*, Xiaodong Xie, Gao Wen
National Engineering Laboratory for Video Technology, Peking University ,Beijing, China

Abstract—**In this paper, we propose a novel interpolation algorithm for adapting the human visual system (HVS) and applying in real-time image upscaling. The defined statistical features are first computed in a local window of the low-resolution (LR) counterpart. Then the most correlated neighbours of a missing pixel in high-resolution (HR) image are adaptively selected based on local structural analysis for the prominent edges and fine textures. Finally, the unknown pixel values are estimated through a designed directional clustering model (DCM) which incorporates HVS information into the weighted coefficients. The extensive experimental results show that the proposed image interpolation method can accurately reconstruct the structures of HR image in term of arbitrary magnification factors and effectively suppress the jaggy/ringing artifacts with low computation complexity.**

Index Terms—**Image interpolation, gradient features, Gaussian model, HVS information**

I. INTRODUCTION

Image interpolation is a technique of producing a HR image from its single-input LR counterpart. This fundamental operation has wide applications in digital photograph resizing, image/video coding and high-definition display. The image enlargement is necessary to recover sharp edges and textures from aliased intensity data. In addition, the interpolation for videos requires to maintain the natural temporal coherence among the consecutive frames.

The classical interpolation methods such as bilinear and bicubic are based on the linear filters or polynomial functions. These simple methods can interpolate quickly and however they often produce undesired artifacts on edges and textures. To preserve sharp structures, some adaptive-edge nonlinear interpolation methods have been developed [1-5]. Gao *et al.* [1] propose to estimate the HR image via combining both the local autoregressive model and the nonlocal adaptive 3-D sparse model as regularized constraints. Giachetti *et al.* [2] propose to use local second order directional information and perform an iterative correction to obtain the interpolated pixels in real-time. Jing *et al.* [4] first estimate the gradients of the HR image based on local similar intensities and then determine the unknown pixels by using their neighbouring LR pixels. To get better visual effects, some works propose to utilize local statistical properties and interpolate unknown pixels by the iterative computation [6-9]. For example, Wei *et al.* [6] propose to generate four binary contrast-guided maps and then conduct different interpolation process according to edge and non-edge pixels. Dai *et al.* [13] design a metric of the soft edge smoothness as a prior and then reconstruct HR images by minimizing the total lengths of all level lines. In addition, the learning-based interpolation methods have been proposed to recover lost high frequency details in LR images from a set of training images [10]. Wavelet transform based methods have also been developed for predicting HR images from decomposed images with the different scales [11].

*The corresponding author. E-mail: hzjia@pku.edu.cn

Although the recent research on image/video interpolation has been made great progress, it still has many challenges to enhance subject quality and reduce the computation cost simultaneously. In order to solve these issues, we propose a novel interpolation framework to faithfully infer the finer image content while preventing visual artifacts by considering the human perceptual characteristics. The proposed method can accommodate arbitrary factors with short processing time. According to the spatial distribution of pixel intensity values along different edge or ridge directions, a set of image features are designed to localize the homogeneous regions or the isophotes in LR images. These features are adopted to estimate the correlation between a missing pixels in HR image with its known neighboring pixels. HVS information is further used to exclude the outliers. Then the neighboring pixels with strong strengths are selected to predict the missing pixels. In order to reducing observable artifacts, a predictive model named DCM is constructed by introducing the distance weights and HVS information. The high-quality upsampling contents can be retrieved while typical artifacts are minimal.

The rest of the organization of this paper is as follows. Section Ⅱ describes the formulation and theory of our image interpolation algorithm. The experimental results are given in Section Ⅲ. Section Ⅳ concludes this paper.

II. DCM-BASED INTERPOLATION ALGORITHM

Our interpolation framework has two main steps including the accurate selection of neighbours and the robust prediction of pixel intensities. The detailed procedure of our algorithm is described in this section.

A. Selection of Neighboring Pixels

The edges and textures with large intensity variations are very important to visual quality. The pixels around the different types of sharp curvatures generally correspond to large gradient magnitudes while the pixels in smooth regions or along the edges have relatively small magnitudes. For the HR image we want to interpolate, its contents need to be consistent with those in the LR image. The unknown pixels at higher resolutions depend on the spatial distribution and the total intensity variations which need to be estimated from the LR image. As shown in the Fig.1, the local structures in red boxes have different distribution types. For the edges in box 1, 3 and 4, the neighboring pixels need be classified in three possible regions including the white, the black and the closeness to the edges. The pixel selection across the edges will cause the blurring effect. For the small and continuous intensity variations in box 2 and 5, the pixels selected along the circle are more reasonable. The pattern in box 6 is complex and the pixel selection along the edges may include the outliers. By analyzing the different selection types, the gradient-based statistical features can be designed.

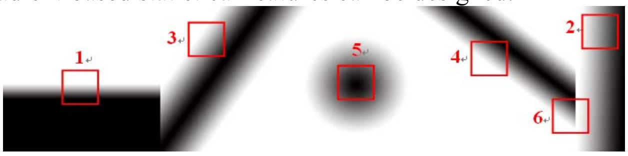


Fig. 1 Different local structures illustration

Without any loss of generality, the size of each local window in the HR image is set as 7×7 pixels. The spatial configuration of the local window is shown in Fig. 2. The black dots represent the known pixels in LR images and the white diamonds represent the unknown pixels in HR images. Based on the extraction from edge metrics, the correct structures can be reconstructed in the new HR image [7].

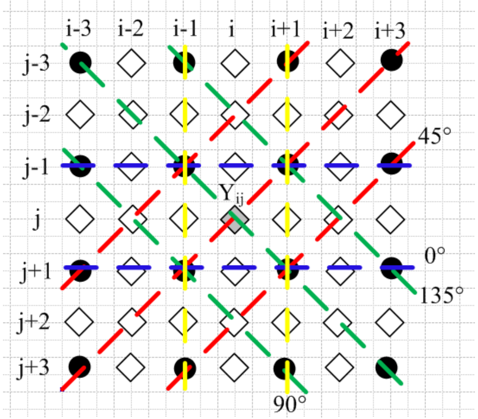


Fig. 2 Spatial configuration in the HR image

The gradient-based features are defined as following:

$$g^{45}=\sum_{m=3,\pm1}\sum_{n=3,\pm1}\left|Y(i+m,j-n)-Y(i+m-2,j-n+2)\right|+s^{45} \tag{1}$$

$$s^{45}=Max\{Y(i+m,j-n)\big|_{m=\pm3,\pm1}^{n=\mp3,\mp1}\}-Min\{Y(i+m,j-n)\big|_{m=\pm3,\pm1}^{n=\mp3,\mp1}\} \tag{2}$$

$$g^{01}=\sum_{m=3,\pm1}\sum_{n=-1}\left|Y(i+m,j+n)-Y(i+m-2,j+n)\right| \tag{3}$$

$$g^{02}=\sum_{m=3,\pm1}\sum_{n=1}\left|Y(i+m,j+n)-Y(i+m-2,j+n)\right| \tag{4}$$

$$g^{0}=(g^{01}+g^{02})/2+s^{0} \tag{5}$$

$$s^{0}=Max\{Y(i+m,j-n)\big|_{m=\pm3,\pm1}^{n=\mp1}\}-Min\{Y(i+m,j-n)\big|_{m=\pm3,\pm1}^{n=\mp1}\} \tag{6}$$

The features along 135° direction g^{135} are similar to g^{45}. The features along 90° direction g^{90} is similar to g^{0}. The pixels along the direction θ in which the intensity variations g^{θ} has the minimum sum will be selected.

$$g^{\theta}=Min\{g^{0},\ g^{90}\ \ g^{45},\ g^{135}\} \tag{7}$$

If the value in the direction 0° or 90° is minimum, all the eight pixels are selected.

B. *Prediction based on DCM*

The selected pixels in the local window are considered as the strong relation with the unknown pixels. The intensity variations are small between this group of neighboring pixels. In order to accurately estimate the interpolated pixels, HVS information and gradient distribution in smooth regions are used to constructed DCM.

Many research has been done on discovering the property of HVS. It has been found that the perception of HVS is more sensitive to luminance contract rather uniform brightness. The ability of human eyes to perceive the magnitude difference between an object and its background depends on the average value of background luminance [12]. This relationship can be used to limit the estimated intensities in reasonable scope (shown in Fig. 3). The threshold is calculated by

$$A_k=\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}I_k(i,j) \tag{8}$$

$$T_k=20.66e^{-0.03A_k}+e^{0.008A_k} \tag{9}$$

where A_k is the mean value of selected pixels and T_k is the maximum visibility threshold.

Based on equation (8) and (9), the pixels selected in the first stage can be further determined the correlation by

If $|Y_i-A_k|>T_k$, Y_i is seen as an outlier;

If $|Y_i-A_k|<T_k$, Y_i is a neighboring pixel;

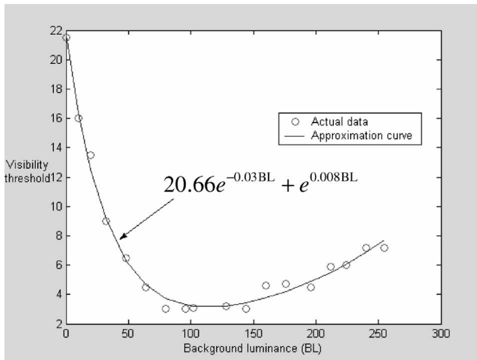


Fig. 1 Visibility thresholds in different background luminance

The gradient field in the smooth LR regions may not correspond to a similarly sloped and spatial distribution at the higher resolution. To avoid the ringing effect and abrupt intensity change, the gradient distributions in smooth edge regions are adopted as weight values in DCM. For different smooth areas, the relationship between the gradient magnitude and the pixel distance can be used to estimate the probability density by training a lot of natural images [8]. A fitting results are shown in Fig. 4. According to the characteristics of the pixel selection, the relationship between the weight and the distance is formulated as a Gaussian model:

$$w_i=\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-(Y_i-c)}{2\sigma^2}} \tag{10}$$

where w_i is weight value, Y_i is the selected pixel, c is the mean value and σ is the variance.

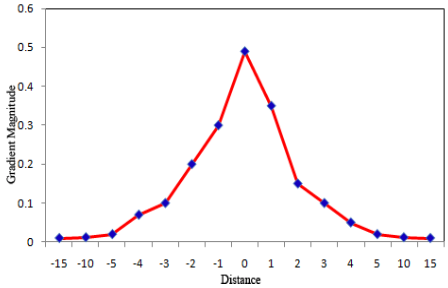


Fig. 4 Fitting results of gradient with different shapes

Based on the equation (9) and (10), the proposed DCM is formulated as the following:

$$\widehat{X} = \frac{\sum_{i=0}^{s} w_i Y_i}{\sum_{k=0}^{s} w_i} \quad (11)$$

Using this model, the unknown $\widehat{X}$ can be estimated accurately.

C. Algorithm In Details

After selecting the strong related neighboring pixels and constructing the predictive functions, the whole framework of interpolation can be implemented as following description:

Input:

The LR image **Y**, the resizing ratio r_x and r_y along the column and row.

Initialization:

Set the size of local detection window d.

Generally, the size is set no less than 3×3 pixels and no more than 12×12 pixels. The size of 6×6 can achieve good results according to a lot of testing experiments.

Loop:

Iterate on the position of the pixels P_{ij}.

i = 1 to i = N (The column number of HR imge).

j = 1 to j = M (The row number of HR image).

Transform the pixels in LR image into the corresponding locations in HR image.

If the distances between the pixels and original point are less than the window size, the unknown pixels are interpolated by using simple bilinear method.

If the magnification factor is integer, the ratio 2× or 3× image is first obtained and then target HR image is continued to upscale by same ratio each time.

If the magnification factor is non-integer, the image with integer magnification is first obtained and then the remainder of unknown pixels are computed.

Select the neighboring pixels from local window.

Compute the predictive pixel values.

Output: the interpolated HR image X.

III. EXPERIMENTAL RESULTS

To fully evaluate the proposed scheme, the comparisons with typical methods including bicubic, SAI [3], Li [5] and Yan [8] are made on subjective visual effect, objective quality and computation time. In order to cover the wide range, eight images are selected as test sets (see Fig.5). The colorful LR images are first transformed RGB space into YUV space and the interpolation is performed only on the luminance channel images. The un-optimized programming of our method is implemented in Matlab 2010. The experiments are executed on a computer of Pentium Core i5 CPU with 2GB memory.

The PSNR results of all methods are computed to evaluate the objective quality. As seen from the results in Table Ⅰ, our method labelled 2 is more robust than other methods. The results also show that the values of PSNR using HVS information (labelled 1) are higher than those without using HVS information (labelled 2). The bicubic method has the fastest computation speed and our method has the second speed due to the detection of local structure (listed Table □). The computation time of our algorithm can be real-time by parallel computation which can utilize multi-threads or GPU acceleration. The subjective comparison results are shown in Fig.6 and Fig.7. Among all methods, our method can achieve best quality with high contrast and minimal artifacts.

TABLE I

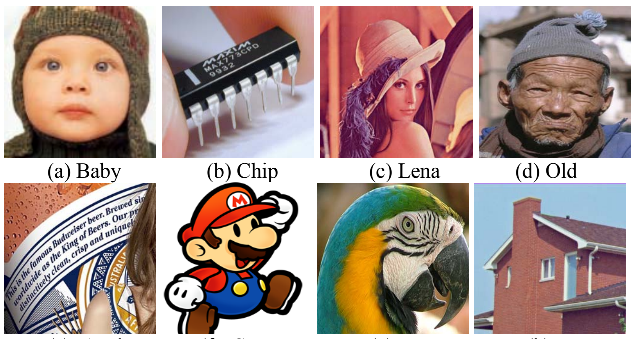
PSNR(DB) RESULTS OF DIFFERENT METHODS WITH FACTOR OF 2

Sample	Bicubic	SAI	Yan	Our[1]	Our[2]
Baby	30.16	31.05	32.08	31.01	**31.64**
Chip	29.87	30.61	31.87	30.93	**32.01**
Lena	29.92	30.28	31.77	30.24	**30.97**
Old	31.02	31.45	31.81	31.19	**31.89**
Anchor	29.34	29.41	29.76	29.89	**30.25**
Cartoon	32.79	33.41	33.55	33.27	**33.65**
Parrot	31.41	31.87	31.78	31.65	**32.14**
House	28.55	29.11	30.01	29.29	**29.53**

TABLE □

COMPUTATION TIME (SECOND) OF DIFFERENT METHODS WITH FACTOR OF 2

Size	Bicubic	SAI	Yan	Our
Baby(128×128)	0.09	1.82	1.69	**0.27**
Chip(244×200)	0.13	4.98	4.97	**0.32**
Lena(220×220)	0.12	4.61	5.51	**0.41**
Old(200×266)	0.11	3.78	5.32	**0.33**
Anchor(600×540)	1.82	70.90	37.81	**3.45**
Cartoon(210×240)	0.14	4.12	5.62	**0.42**
Parrot(310×250)	0.22	8.34	8.29	**0.68**
House(256×256)	0.20	8.10	8.07	**0.61**



(a) Baby (b) Chip (c) Lena (d) Old

(e) Anchor (f) Cartoon (g) Parrot (h) House

Fig. 5 The testing image sets

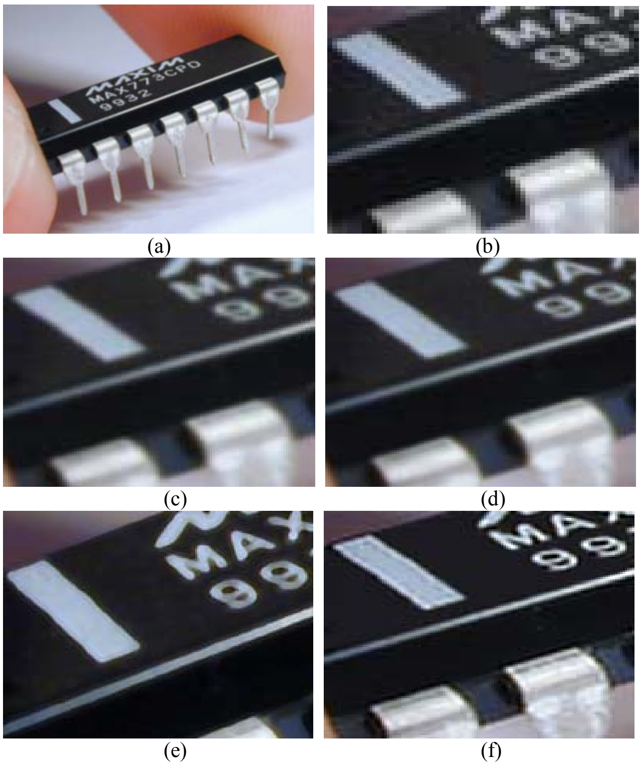


Fig. 6 Image upsampling with the factor of 2. (a) Input LR image. (b) The corresponding patch image. (c) The bicubic results. (d) The SAI results. (e) Yan's results. (f) Our results

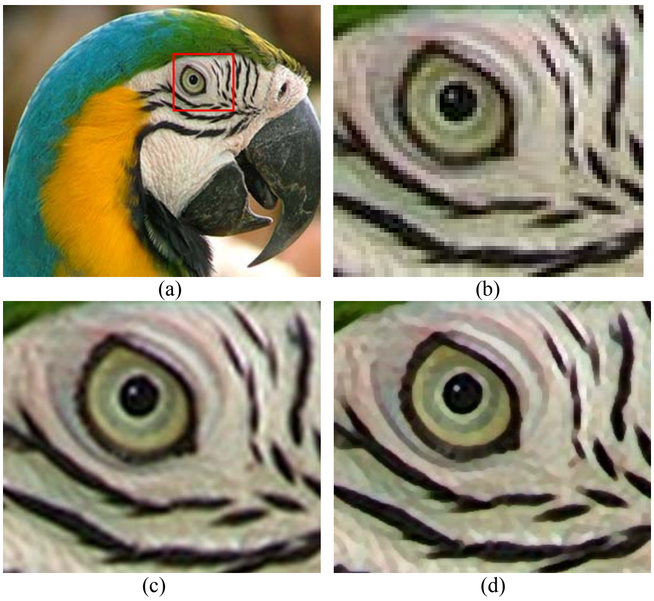


Fig. 7 Image upsampling with the factor of 3. (a) Input LR image. (b) The bicubic results. (c) Li's results. (d) Our results

IV. CONCLUSIONS

This paper presents an effective image interpolation scheme, which is formulated by using noniterative HVS-guided DCM. Based on the structural similarity between images at different resolutions, the local boundaries with large intensity variations in LR image are fast detected. The total gradient variations and standard deviation along every possible orientation are defined as the metrics. By applying these statistical features to the proposed decision function, the outliers along the edges can be reliably excluded and the neighbors of the unknown pixels in local windows can be correctly determined. Due to strong structure-preserving capacity, this selection stage can avoid the jaggy and blurring. HVS properties are used to exclude the outliers and increase the contrast. According to gradient distributions, a predictive model named DCM is constructed to estimate the pixel intensities which are consistent with the human perception. This predictive stage can avoid the ringing. Finally, the whole interpolated image can be obtained by pixel-wise processing. The experimental results show that our method can achieve performance improvements over the state-of-the-art methods in visual quality. This scheme can be exploited as varying-ratio image and video upscaler in real-time.

ACKNOWLEDGMENT

This work is supported by the Chinese National Natural Science Foundation under contract No.61171139, National Key Technology Research and Development Program of the Ministry of Science and Technology of China under contract No.2014BAK10B00, Major National Scientific Instrument and Equipment Development Project of China under contract No.2013YQ030967 and Beida(Binhai) Information Research.

REFERENCES

[1] X.W. Gao, J. Zhang, F. Jiang, and X.P. Fan, "High-quality image interpolation via local autoregressive and nonlocal 3-D sparse regularization," in *Proc. IEEE Conf. Visual Comm. and Image Process.*, Nov. 2012, pp. 1–6.

[2] A. Giachetti and N. Asuni, "Real-time artifact-free image upscaling," *IEEE Trans. Image Process.*, vol. 20, no. 10, pp. 2760–2768, Oct. 2011.

[3] X. J. Zhang and X. L. Wu, "Image interpolation by adaptive 2-D autoregressive modeling and soft-decision estimation," *IEEE Trans. Image Process.*, vol. 17, no. 6, pp. 887–896, Jun. 2008.

[4] G. M. Jing, Y. K. Choi, J. Y. Wang and W. P. Wang, "Gradient guided image interpolation," in *Proc. IEEE Conf. Image Process.*, Oct. 2014, pp. 1822–1826.

[5] M. D. Li, J. Y. Liu, J. Ren and Z. M. Guo, "Adaptive general scale interpolation based on weighted autoregressive models," *IEEE Trans. Syst. Video Technol.*, vol. 25, no. 2, pp. 200–211, Feb. 2015.

[6] Z. Wei and K. K. Ma, "Contrast-guided image interpolation," *IEEE Image Process.*, vol. 22, no. 11, pp. 4271–4285, 2013.

[7] R. Fattal, "Image upsampling via imposed statistics," *ACM Trans. Graphics*, vol. 26, no. 3, pp. 1–8, Jul. 2007.

[8] Q. Yan, Y. Xu, X.K. Yang and T. Nguyen, "Single image super-resolution based on gradient profile sharpness," *IEEE Image Process.*, no. 99, pp. 1–18, 2015.

[9] J. Y. Wen and Y. Wang, "An efficient framework for image interpolation using weighted surface approximation," in *Proc. IEEE Conf. Visual Comm. and Image Process.*, Dec. 2014, pp. 354–357.

[10] W.T. Freeman, T.R. Jones and E.C. Pasztor, "Example-based super-resolution," *IEEE Comput. Image Graph. Appl.*, vol. 22, no. 2, pp. 56–65, Mar. 2002.

[11] S.G. Chang, Z. Cvetkovic and M. Vetterli, "Locally adaptive wavelet-based image interpolation," *IEEE Image Process.*, vol. 15, no. 6, pp. 1471–1485, Jun. 2006.

[12] C. T. Lin, K. W. Fan, H. C. Pu, and S. M. Lu, "An HVS-directed neural-network-based image resolution enhancement scheme for image resizing," *IEEE Trans. Fuzzy Systems.*, vol. 15, no. 4, pp. 605–615, Aug. 2007.

[13] S. Y. Dai, M. Han and W. Xu, "Softcuts: a soft edge smoothness prior for color image super-resolution," *IEEE Image Process.*, vol. 18, no. 5, pp. 969–981, 2009.