International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.8, No.10 (2015), pp.97-106
http://dx.doi.org/10.14257/ijsip.2015.8.10.11

Image Interpolation Based on Saliency Information

Lei Shu, Yuming Fang, Zhijun Fang, Yong Yang, Feiniu Yuan and Fengchang Fei

School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China 330032
sleker@163.com

Abstract

In this paper, we propose an image interpolation algorithm based on the characteristics of the Human Visual System (HVS). Firstly, the saliency map is extracted for the input image by the feature contrast. Based on the saliency detection algorithm, the input image can be divided into salient regions and non-salient regions. A non-local means based interpolation and the bicubic interpolation operations are adopted to implement the image interpolation for salient and non-salient regions in visual scenes, respectively. Experimental results show that the proposed method improves the perceptual quality of the salient regions during image interpolation. The PSNR and SSIM results demonstrate that the proposed method can obtain better performance than existing related image interpolation methods.

Keywords: *image super-resolution, image interpolation, saliency map, human visual system, saliency detection*

1. Introduction

Single image super-resolution (SISR), referring to the technique of generating a high-resolution (HR) image from a low-resolution (LR) one, is meaningful for visual surveillance, high-definition TV, medical image processing, *etc.* Currently, super-resolution image reconstruction method can be roughly divided into four categories [1]: frequency domain based, interpolation based, regularization based and learning based methods.

The frequency domain based method is firstly to be presented in [2], where the SR computation for the noise-free low-resolution images is taken into consideration. It transforms the low-resolution image into the Discrete Fourier transform (DFT) domain and combines them into a final result according to the relationship between the aliased DFT coefficients of the low-resolution images. Then the processed DFT coefficients are transformed back to the spatial domain to obtain a high-resolution image. Later, Discrete cosine transform (DCT) is exploited to perform fast image de-convolution for SR image computation [3].

It is well accepted that the wavelet transform is a powerful and efficient multi-scale representation tool of the image. And many studies use the wavelet transform to address the SR problem to recover the high-frequency information lost or degraded during the image acquisition process. The wavelet transform based image interpolation approaches [4-5] usually treat the low-resolution image as the low-pass filtered subbands of the unknown wavelet-transformed high-resolution image, which is used to estimate the finer scale subband coefficients. Then the inverse wavelet transform is applied to produce the high-resolution image. The frequency domain based method is intuitive and simple, but it is difficult to deal with the problem of noise and add priori information in the process [6]. Temizel *et al.* [7] proposed a wavelet based image interpolation method, which implies that the magnitudes of wavelet coefficients corresponding to the same spatial location tend to propagate from lower scales to higher resolution scales. Inter-scale dependency

ISSN: 2005-4254 IJSIP

based image interpolation method in the wavelet domain is proposed in the study [8]. The method employs a Gaussian Mixture Model (GMM) to estimate the magnitude of wavelet coefficients, and get the parameters of the GMM from subbands. Chang *et al.* used a wavelet transform to extract sharp information in the low-resolution image and then applies interpolation which adapts to the image local smoothness or singularity characteristics [9]. The algorithm yields sharper images with high computational complexity.

The interpolation based method processes information with a target in the image plane as grid fitting or interpolation. Image interpolation is usually accomplished by applying successively 1D kernel interpolation on horizontal and vertical directions. The classical image interpolation algorithms, such as the nearest neighbor interpolation [10], bilinear interpolation [10], bi-cubic convolution [11] and bi-cubic spline interpolation [12], are with low computational complexity and can obtain promising performance in the smooth regions of the image. However, these traditional methods inherently assume smoothness constraints on the signal and, as a result, they typically generate block, saw-tooth, and other issues occurring in the edges.

In order to enhance the visual effect of edges, many interpolation algorithms have been proposed based on edge detection [13-14]. The main idea of these algorithms is to interpolate pixels along the direction of the edge rather than the region across the edge. A new edge-directed interpolation (NEDI) algorithm is designed in [13] to try to improve edge pixel interpolation results However, it requires much matrix inversion and thus is time-consuming. Zhou *et al.* present a colorful image interpolation method [14], where the pixel value is determined by the strength of the edge. This method is simple and fast, but it only detects the edge along diagonal which might degrade the reconstruction of visual effect.

To try to get better performance, some studies propose image super-resolution algorithms based on regularization [15-17]. The basic idea of these methods is to add priori knowledge of high resolution image. The maximum a posterior (MAP) [18] is one of the most popular regularization based algorithms for image interpolation. The regularization term plays a critical role in controlling the final result of image quality. The key issue of such methods is how to design effective regularization term and select the optimal regularization coefficient. The regularization term is used to determine the reconstructed high-resolution images. By regarding the image as a locally smooth data field, the Markov random field (MRF) [19] is adopted as a prior image model. To address the problem of automatically determining the optimal regularization coefficient, some works have been developed during the process of estimating the unknown high-resolution image [20-21], rather than applying an experimentally fixed value.

Recently, the learning based super-resolution reconstruction methods have attracted much attention in the research community. Machine learning techniques are used to map the relationship between high frequency information and low frequency information in the training data. The missing high frequency information of the test image can be predicted to improve the performance of resulted high resolution image [6]. A super-resolution method by sparse coding is proposed in the study [22] and has been widely recognized [23-24]. The basic idea is to seek a sparse representation for each patch of the low-resolution image and use this representation to generate the high-resolution image. Two training dictionaries of the low-resolution and high-resolution image patches are established for the learning process. The sparse representation of a low-resolution image patch can be applied with the high-resolution image patch dictionary to generate the high-resolution image patches.

The problem with the sparse coding based image interpolation method is that they are required to establish over-complete dictionaries and the assumption of these methods is that the input image patches are linearly sparse representation in the training set. For natural images, the establishment of over-complete dictionaries is complicated, and it suffers from the problem of potential instability, likely to produce visual artifacts. Meanwhile, the learning based super-resolution algorithm is usually time-consuming, and requires a lot of training data. Thus, it is suitable for image pre-processing system offline.

Although there are various image super-resolution methods introduced above, they rarely take into account the characteristics of the Human Visual System (HVS). Generally, the visual information entering human eyes is much more than that handled by the central nervous system. It is hard to identify the objects and their relationships immediately for human eyes when viewing a natural scene. Therefore, most processing resources will be allocated to the saliency regions by human brain, rather than a uniform distribution to all regions of the image [25].

Based on our previous work [26], we propose a novel image interpolation method based on saliency information. Firstly, the saliency map of low-resolution image is extracted by the saliency detection model [26]. Then different interpolation methods are applied to salient regions and non-salient regions to obtain the final interpolated image. Experimental results show that the proposed method can obtain better perceived quality than existing related interpolation methods.

The remainder of the paper is organized as follows. In Section II, the details of our proposed method are provided to better understand the advantages of the proposed framework. Section III presents experimental results and Section IV concludes.

2. The Proposed Method

Visual attention is an important mechanism to process visual information in the HVS. When observers look at a natural scene, they will focus on the salient regions while ignoring other non-salient regions. So we believe that the performance of reconstruction image in salient regions will significantly affect subjective image quality assessment. In this paper, we present a novel interpolation method by using the saliency information and non-local means interpolation approach, aiming to improve the interpolation quality in salient regions. The framework is illustrated in Fig. 1. The steps of the proposed method are as follows.

1. Calculate the saliency map of LR image by using the saliency detection model;
2. Apply bicubic interpolation operation to non-salient regions for image interpolation;
3. Apply non-local means interpolation operation to salient regions for image interpolation;
4. Combine the interpolation results from bicubic and non-local means based interpolation methods to get final HR image.

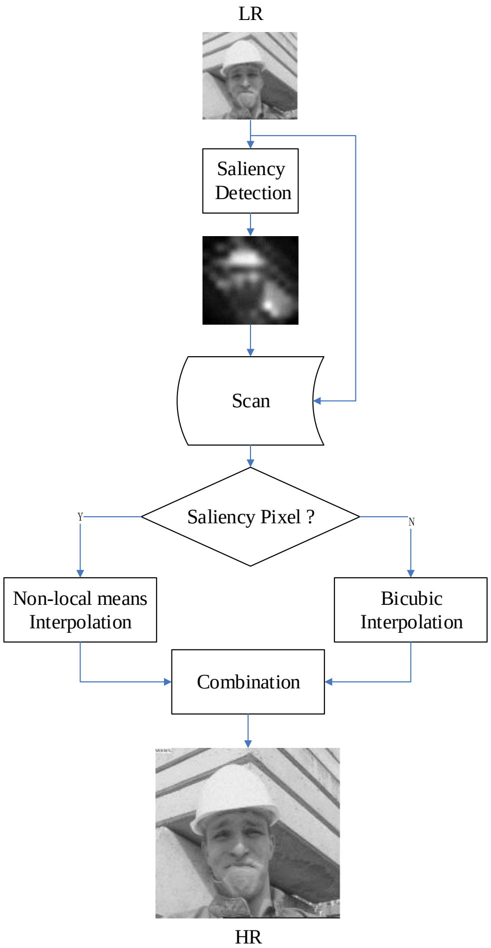

Figure 1. The Framework of the Proposed Method

2.1. Saliency Detection Model

Our previous study of saliency detection first divides the input images into small image patches and then calculates the saliency value of each image patch by computing the differences between the quaternion Fourier transform (QFT) amplitude spectrum of this patch and all other patches in the image [26]. The weights of these differences are determined by the visual impacts of the human visual sensitivity. The saliency value of image patch i can be expressed as follows:

$$S_i = \sum_{j \neq i} w_{ij} d(i, j) \tag{1}$$

where $d(i, j)$ represents the difference between image patches i and j, w_{ij} is the weight for the patch difference between image patches i and j.

In this model, the color and intensity channels are used in QFT to get the amplitude spectrum for each image patch, which is exploited to compute the differences between image patches. According to the related property of the human primary visual cortex [27], the red-green and blue-yellow double opponency of the LR image are calculated for feature extraction. As shown in the study [26], the feature differences between image patches can be represented by the differences of amplitude spectrum between image patches.

The weighting parameters w_{ij} in (1) can be calculated as follows [26]:

$$w_{ij} = \frac{1}{C_0 \exp(af \frac{e + e_2}{e_2})} \tag{2}$$

where f is the spatial frequency (cycles/degree), e is the retinal eccentricity (degree); C_0 is the minimum contrast threshold; a is the spatial frequency decay constant; e_2 is the half-resolution eccentricity. According to the experiments reported in [29], these parameters are set to $C_0 = 1/64$; $a = 0.106$; and $e_2 = 2.3$.

Based on the Eq. (1), the saliency value for the image patch i is represented as all the contributions from the patch differences between the image patch i and all other image patches in the image.

2.2. Bicubic Interpolation

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

$$Y(u,v) = \sum_i \sum_j y(i,j) h\left(\frac{u-u_k}{\Delta u}\right) h\left(\frac{v-v_k}{\Delta v}\right) \tag{3}$$

where $(\Delta u, \Delta v)$ are sampling intervals, $h()$ is the interpolation kernel and $y(i,j)$ represents the pixel array in the LR grid. The HR signal is obtained by resampling (6) on a finer grid. In [8], the cubic convolution kernel is given as,

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

2.3. Non-local Means Interpolation

In the study [30], Protter *et al.* present a non-local means based image super-resolution method and achieve good results. However, this method needs to process video sequences with general motion patterns, so it is time-consuming and unsuitable for single image. Thus, we modify that method to use it for the salient regions of images. The non-local means interpolation is robust to noise and preserves the local characteristics of the image. The algorithm is depicted as follows.

Input:

y - the salient region of LR

s - scaling factor

q - the size of the low resolution image (R^{L})

p - the size of the high resolution image (R^{H}), $p = s(q-1)+1$

Y - an initial estimation of salient region in the high-resolution image

Initialization:

Set $\mathrm{X}=\mathrm{Y}$

Set V and W to be zero, and the same size as Y

Steps:

For $(k,l) \in \mathrm{Y}$ and $(i,j) \in \mathrm{y}$, where $(si,sj) \in Neighbor(k,l)$.

1) Compute the weight:

$$w(k,l,i,j) = \exp(-\|\hat{\mathrm{R}}_{k,l}\mathrm{Z} - \hat{\mathrm{R}}_{si,sj}\mathrm{Y}\|_2^2 / 2\sigma^2)$$

$\|\hat{\mathrm{R}}_{k,l}\mathrm{Z} - \hat{\mathrm{R}}_{si,sj}\mathrm{Y}\|_2^2$ is the Euclidean distance between the patches (k,l) and (si,sj)

2) Accumulate Inputs:

Extract the low-resolution patch $\mathrm{R}_{i,j}^{L}\mathrm{y}$

Upscale it by zero-filling

Accumulate it in its proper location

$$\mathrm{V} = \mathrm{V} + w(k,l,i,j)(\mathrm{R}^{H}_{k,l})^{T}\mathrm{D}^{T}_{p}\mathrm{R}^{L}_{i,j}\mathrm{y}$$

D^{T}_{p} is the zero-filling operator

3) Accumulate Weights:

$$\mathrm{W} = \mathrm{W} + w(k,l,i,j)(\mathrm{R}^{H}_{k,l})^{T}\mathrm{D}^{T}_{p}\mathrm{R}^{L}_{i,j}\mathrm{E}$$

4) Normalization:

$$\mathrm{X}(k,l) = \mathrm{V}(k,l) / \mathrm{W}(k,l)$$

Output:

High-resolution saliency region X

3. Experimental Results

In this section, we have conducted the experiment to demonstrate the performance of the proposed method. Three video sequences are used to conduct the comparison experiment: Foreman, Miss America and Suzie from [31], are also used in the study [16]. The frames from each video are shown in Fig.3. Six different sample frames from each video are used to compare the proposed method with other four popular interpolation algorithms: bilinear interpolation (BI) [7], cubic interpolation (CC) [8], new edge-directed interpolation (NEDI) [10] and iterative curvature based interpolation (ICBI) [32].

The comparison results are shown in Tables 2 and 3. From these tables, it can be seen that the performance of proposed method is the best with the results of both PSNR and SSIM. The most textured image 'Forman', is with the best performance among all test video frames. Fig.3 shows the comparison of portions of the interpolated results from different methods. The proposed method can obtain the best perceptual quality in the salient regions.

Table 2. Comparison of Different Interpolation Methods on PSNR

Image	BI	CC	NEDI	ICBI	Proposed
Foreman	28.5081	29.0793	28.8135	29.1844	29.4103
Miss America	34.2415	34.8554	34.1284	34.6785	35.1025
Suzie	29.6711	30.0323	29.2720	29.8561	30.1647
average	30.8069	31.3223	30.7380	31.2397	31.5592

Figure 3. Video Frame Samples from Different Videos

Top: HR images. Bottom: LR images. From left to right: Foreman; Miss America; Suzie.

Table 3. Comparison of Different Interpolation Methods on SSIM

Image	BI	CC	NEDI	ICBI	Proposed
Foreman	0.8485	0.8595	0.8507	0.8532	0.8687
Miss America	0.8939	0.8988	0.8894	0.8951	0.9014
Suzie	0.8270	0.8337	0.8074	0.8294	0.8389
average	0.8565	0.8640	0.8492	0.8592	0.8697

4. Conclusion

We have proposed a new image interpolation method based on the saliency detection model. Firstly, the saliency detection model is used to calculate the saliency map of low-resolution images. Then Non-local means interpolation operation is applied to salient regions to improve the interpolation quality. Experimental results show that the proposed method can obtain better performance than conventional interpolation methods.

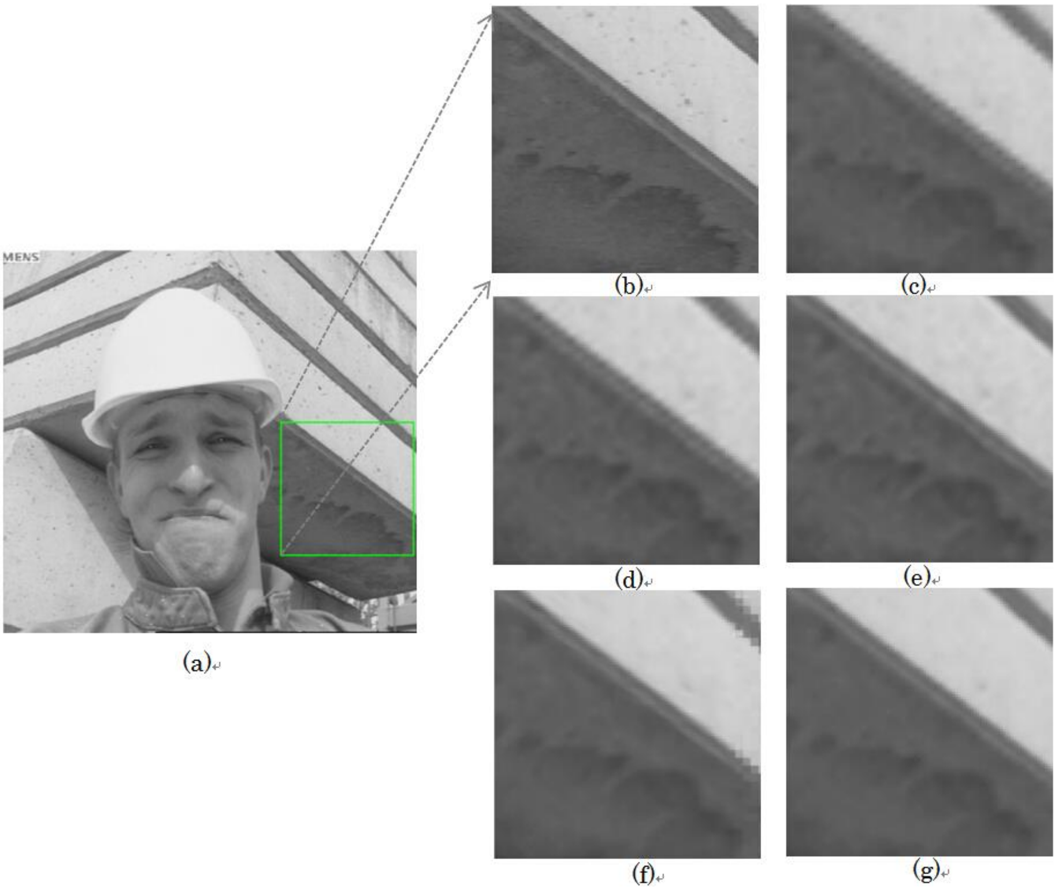

Figure 4. Portions of the interpolated results of the 'Foreman' (a) Original image (b) a small region of the original image (c) The result from BI (d) The result from CC (e) The result from ICBI (f) The result from NEDI (g) The result from the Proposed method.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 61461021, 61462031), the Science and Technology Application Project of Jiangxi province (No. KJLD14031), and the Key Academic Leader Plan in Jiangxi Province (No. 20133BCB22005).

References

[1] T. Jing and M. Kai-Kuang, "A survey on super-resolution imaging". Signal, Image and Video Processing, vol. 5, no. 3, (**2011**), pp. 329-342.

[2] R. Y. Tsai and T. S. Huang, "Multiframe image restoration and registration," In: Huang, T.S. (ed.) Advances in Computer Vision and Image Processing, (**1984**), JAI Press Inc., London.

[3] S. Rhee and M. G. Kang, "Discrete cosine transform based regularized high-resolution image reconstruction algorithm", Opt. Eng. 38, (**1999**), pp. 1348–1356.

[4] H. Ji and C. Fermuller, "Wavelet-based super-resolution reconstruction: theory and algorithm", In: Proceedings of the European Conference on Computer Vision, Graz, Austria, (**2006**), pp. 295–307.

[5] H. Ji and C. Fermuller, "Robust wavelet-based super-resolution reconstruction: theory and algorithm", IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, (**2009**), pp. 649–660.

[6] H. Su, J. Zhou and Z. H. Zhang, "Survey of Super-resolution Image Reconstruction Methods," Acta Atuo-matica Sinica, vol. 39, no. 8, (**2013**), pp. 1202-1213.

[7] A. Temizel and T. Vlachos, "Wavelet domain image resolution enhancement", IEE Proc.-Vis. Image Signal Process., vol. 153, no. 1, (**2006**), pp. 25-29.

[8] H. W. Dong, K. Eom and Y. S. Kim, "Image Interpolation based on inter-scale dependency in wavelet domain", (**2004**), International Conference on Image Processing (ICIP).

[9] S. G. Chang, Z. Cvetkovic and M. Vetterli, "Locally Adaptive Wavelet-Based Image Interpolation". IEEE Transactions on Image Processing, vol. 15, no. 6, (**2006**).

[10] S. S. Rifman, "Digital rectification of ERTS multispectral imagery," In Proceedings of the Symposium on Significant Results Obtained from the Earth Resources Technology Satellite-1, 1(B), (**1973**), pp.1131-1142.

[11] R. G. Keys, "Cubic convolution interpolation for digital image processing," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 29, no. 6, (**1981**), pp. 1153-1160.

[12] H. S. Hou and H. C. Andrews, "Cubic splines for image interpolation and digital filtering," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 26, no. 6, (**1978**), pp. 508-517.

[13] X. Li and M. T. Orchard, "New edge-directed interpolation," IEEE Transactions on Image Processing, vol. 10, no. 10, (**2001**), pp. 1521-1527.

[14] D. W. Zhou and X. L. Shen, "Edge-directed bicubic color image interpolation," Acta Automatica Sinica, vol. 38, no. 4, (**2012**), pp. 525-530.

[15] K. V. Suresh and A. N. Rajagopalan, "Robust and computationally efficient super-resolution algorithm," Opt. Soc. Am. J. A Opt. Image Sci. 24, (**2007**), pp. 984–992.

[16] S. Belekos, N. P. Galatsanos and A. K. Katsaggelos, "Maximum a posteriori video super-resolution using a new multichannel image prior," IEEE Transactions on Image Process, vol. 19, (**2010**), pp. 1451–1464.

[17] H. Shen, L. Zhang, B. Huang and P. Li, "A MAP approach for jointmotion estimation, segmentation, and super resolution," IEEE Transactions on Image Process, vol. 16, (**2007**), pp. 479–490.

[18] C. Bishop, "Bayesian image super-resolution," Advances in Neural Information Processing Systems, (**2002**), pp. 1303–1310.

[19] S. Z. Li, "Markov Random Field Modeling in Computer Vision," Springer, (**1995**), New York.

[20] L. C. Pickup and D. Capel, S. J. Roberts, A. Zisserman, "Bayesian image super-resolution, continued. In: Scholkopf, B., Platt, J., Hoffman, T. (eds.) Advances in Neural Information Processing Systems, (**2006**), pp. 1089–1096, MIT Press, Cambridge.

[21] M. V. W. Zibettia, F. S. V. Baznb and J. Mayer, "Estimation of the parameters in regularized simultaneous super-resolution," Pattern Recognition, Letter, vol. 32, (**2011**), pp. 69–78.

[22] J. Yang, J. Wright, T. S. Huang and Y. Ma, "Image super-resolution as sparse representation of raw image patches," IEEE Conference on Computer Vision and Pattern Recognition, (**2008**).

[23] I. K. Kwang and K. Younghee, "Single-image super-resolution using sparse regression and natural image prior," IEEE Trans Pattern Anal Mach Intell, vol. 32, no. 6, (**2010**), pp. 1127-33.

[24] J. Yang, J. Wright, T. S. Huang and Y. Ma, "Image super-resolution via sparse representation," IEEE Transactions on Image Process, vol. 19, no. 11, (**2010**), pp. 2861-73.

[25] H. Pashler, "The Psychology of Attention", Cambridge, MA, (**1997**), MIT Press.

[26] Y. M. Fang, W. S. Lin, B. S. Lee, C. T. Lau, Z. Z. Chen, C. W. Lin, "Bottom-Up Saliency Detection Model Based on Human Visual Sensitivity and Amplitude Spectrum," IEEE Transactions on Multimedia, vol. 14, no. 1, (**2012**), pp. 187-198.

[27] S. Engel, X. Zhang, and B. Wandell, "Colour tuning in human visual cortexmeasured with functional magnetic resonance imaging," Nature, vol. 388, no. 6, (**1997**), pp. 68–71.

[28] T. Ell and S. Sangwin, "Hypercomplex Fourier transforms of color images," IEEE Trans. Image Process., vol. 16, no. 1, (**2007**), pp. 22–35.

[29] W. S. Geisler and J. S. Perry, "A real-time foveated multi-solution system for low-bandwidth video communication," in Proc. SPIE, vol. 3299, (**1998**) July, pp. 294–305.

[30] M. Protter, M. Elad, H. Takeda and P. Milanfar, "Generalizing the Non-local means to Super-Resolution Reconstruction," IEEE Transactions on Image Process, vol. 18, no. 1, (**2009**), pp. 36-51.

[31] http://www.cs.technion.ac.il/~matanpr/NLM-SR/Experiments/

[32] A. Giachetti and N. Asuni, "Fast artifacts-free image interpolation," In: Proceedings of the British Machine Vision Conference, (**2008**), pp. 123-132, Leeds, UK: BMVC.

Authors

Lei Shu is a PhD candidate in management science and engineering, school of information technology, Jiangxi University of Finance and Economics, Nanchang, China. His research interests include image processing and machine vision.

Yuming Fang is currently a faculty member in the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China. He received the Ph.D. degree in Computer Engineering from Nanyang Technological University, Singapore in 2013. He was a (visiting) Postdoc Research Fellow in IRCCyN lab, PolyTech' Nantes & Univ. Nantes, Nantes, France, University of Waterloo, Waterloo, Canada and Nanyang Technological University, Singapore. His research interests include visual attention modeling, visual quality assessment, image retargeting, computer vision, 3D multimedia processing, etc.

Zhijun Fang received the Ph.D. degree from Shanghai Jiaotong University, Shanghai, China. He is currently a professor in the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China. His research interests include image processing, video coding, and pattern recognition. He was the General Chair of HHME2013 (the 9th Joint Conference on Harmonious Human Machine Environment) and the General Co-Chairs of ISITC2014 (2014 International Symposium on Information Technology Convergence).

Yong Yang received the Ph.D. degree in biomedical engineering from Xi'an Jiaotong University, China, in 2005. From 2009 to 2010, he was a postdoctoral research fellow at Chonbuk National University, Republic of Korea. He is currently a Full Professor of the School of Information Technology, Jiangxi University of Finance and Economics, China. He has been awarded the title of Jiangxi Province Young Scientist since 2012. His current research interests include image processing, medical image processing and analysis, and pattern recognition.

Feiniu Yuan received the B.Eng. and M.E. degrees in mechanical engineering from the Hefei University of Technology, Hefei, China, in 1998 and 2001, respectively, and the Ph.D. degree in pattern recognition and intelligence system from the University of Science and Technology of China (USTC), Hefei, in 2004. From 2004 to 2006, he worked as a post-doctor with USTC. From 2010 to 2012, he was a Senior Research Fellow with Singapore Bioimaging Consortium, Agency for Science, Technology and Research, Singapore. He is currently a professor and a PhD supervisor with Jiangxi University of Finance and Economics. His research interests include 3D modeling, image processing and pattern recognition.

Fengchang Fei is a PhD candidate in management science and engineering, school of information technology, Jiangxi University of Finance and Economics, Nanchang, China. His research interests include video processing and machine vision.