A Spatial Inter-view Auto-regressive Super-resolution Scheme for Multi-view Image via Scene Matching Algorithm

Min Gao[1], Siwei Ma[2],
[1]Department of Computer Science and Technology
Harbin Institute of Technology
Harbin, China
mgao@hit.edu.cn, swma@jdl.ac.cn

Debin Zhao[1], Wen Gao[1,2]
[2]School of Electronics Engineering and Computer Science
Peking University
Beijing, China
{dbzhao,wgao}@jdl.ac.cn

Abstract—**Binocular suppression theory states that the stereo vision quality is not much influenced by asymmetric degradation of the individual views. Based on these findings, mixed resolution (MR) multi-view framework jointly utilizes the lower and full resolution images to reduce the data amount, while maintaining good stereo vision quality. To enhance the resolution of the lower resolution image, a novel super-resolution scheme for the MR multi-view framework is presented in the paper. It is based on the assumption that the image is modeled as a 2D piecewise auto-regressive process. In the scheme, each pixel to be interpolated is estimated as the linear weighted summation of the pixels, which are consisted of the spatial neighboring ones from lower resolution image in the current view and the ones from the full resolution image in the neighboring views. To get the corresponding pixels in the neighboring views that match the scene to be reconstructed, a window based scene-matching approach is used. Through exploiting the spatial correlation and the inter-view correlation, the proposed scheme achieves a significant gain in PSNR and visual quality for the test sequences.**

I. INTRODUCTION

With the rapid development of the techniques on the multi-view acquisition and display, multi-view video is becoming one of the most promising applications in video entertainment, such as free view point video [1], 3DTV and immersive teleconference. However, there are still some challenges in the multi-view applications. Multi-view compression is one of the challenges, as the data size of the multi-view video tremendously increases with the increment of the number of views. Nevertheless, recent achievement in human visual system suggests that attractive data amount reduction can be achieved by reducing the resolution of the images in multi-view videos while maintaining good stereo vision quality.

The psycho-visual studies [2] in stereoscopic vision states that the stereo vision quality is not affected when one of the views is low-filter passed. It has inspired a lot of multi-view image acquisition and compression techniques, such as mixed resolution (MR) framework. In MR framework, one view image is acquired (or compressed) at full resolution, whereas the other view image is degraded by lower spatial resolution. Several MR coding frameworks for mobile devices have been proposed in [3] [4]. The MR framework integrating the temporal scalability has been investigated in [5]. A prediction method for macro-blocks in lower resolution views are proposed in [6]. A method for high resolution synthesis of lower resolution video at the decoder side was proposed in [7]. The objective gains of the MR frameworks under different down-sampling ratios at low bit-rate have been reported in [8].

Although the MR framework for multi-view video can reduce the data amount, it also has some flaws. First, a full resolution video is used in many scenarios. Second, the user will feel uncomfortable if low quality image is presented continuously to one eye.

To enhance the lower resolution image in MR framework, we can use the traditional image interpolation algorithms [9-13]. However, these algorithms neglect the available information from the full resolution image in neighboring views. It was proposed to utilize the high frequency information from neighboring views through the depth-image-based rendering (DIBR) in [14]. The method requires a true depth in advance. In [15], a dual regularization approach is proposed by exploiting both low and full resolution images. The method needs to iteratively reconstruct the high resolution image and perform the disparity estimation at each iteration step. However, the depth map is not always available and the disparity estimating process is time consuming.

A spatial inter-view autoregressive approach is proposed in the paper to exploit the spatial and inter-view correlation. The proposed method utilizes a scene matching algorithm to find the corresponding block in the full resolution image, and then the autoregressive model is used to estimate the missing pixels.

The paper is organized as follows; section 2 presents an introduction to MR multi-view framework; the proposed super-resolution algorithm is presented in section 3; experimental results are illustrated in section 4 and section 5 concludes the paper.

978-1-4673-5762-3/13/$31.00 ©2013 IEEE

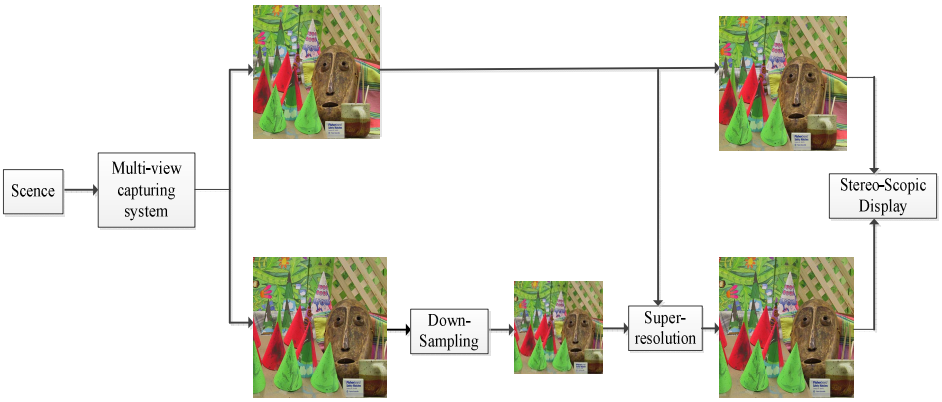


Fig.1 Mixed resolution (MR) multi-view framework

II. Overview of Mixed Resolution Multi-view Framework

The mixed resolution (MR) framework is proposed to reduce the data amount to be stored. As illustrated in Fig.1, the multi-view images captured by the cameras are considered as full resolution ground truth. To reduce the bits required to store these images, the image from one view is down-sampled to lower resolution version. When the multi-view images are displayed, the lower resolution image is up-sampled to the original resolution version and then displayed by the stereoscopic display device.

The characteristic of MR multi-view framework makes the super-resolution scheme has some different aspects compared with the conventional ones. First, the images in MR framework have different resolutions, while the images used in conventional image super-resolution methods are all lower-resolution version. Second, the super-resolution method in MR multi-view framework can enhance the resolution of the lower-resolution image by using the high resolution one, which is available in the neighboring views and can provide more detail information.

III. Spatial Inter-view Autoregressive Super-resolution for Multi-view Image in MR Framework

A. Overview of the Proposed Super-resolution Scheme

In the proposed super-resolution scheme, the missing pixels are interpolated in two passes. The work of the two passes is shown in Fig.2.

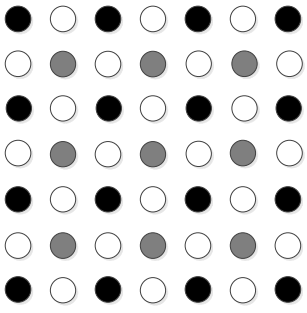


Fig.2 Formation of the interpolation through two passes

In Fig.2, the solid dots are the original pixels in LR image, the shaded pixels are the ones to be interpolated in the first pass, and the empty pixels are the interpolated pixels in the second pass. So the pixels interpolated in the first pass can be used as the reference pixels in the second pass.

In the mixed resolution framework for multi-view setup, we assume that the LR image is from view n, and image from its neighboring view k is full resolution.

Since the image from the neighboring view is full resolution, which can provide more structural information, more reference pixels from neighboring view are used to make full use of this information. The pixel set S_k from the neighboring view k obtained through the scene-matching algorithm consists of 9 pixels in the center, diagonal and axial directions, as shown in equation (1); in addition to the inter-view correlation, the spatial correlation is also exploited in the proposed method. In the first pass, the spatial neighboring pixel set $S_\times$ is used, which consists of 4 neighboring pixels in the diagonal directions, as shown in equation (2); in the second pass, the 4 neighboring pixels in the axial directions is used, which are referenced as S_+, as shown in equation (3). The formation of S_k, $S_\times$ and S_+ is illustrated in Fig.3.

So for every missing pixel in LR image from view n, the pixel value is estimated as the weight summation of the pixel values in S_k taken from the neighboring views k and its spatial neighbors $S_\times$ and S_+.

$$S_k(i,j) = \{(i-x, j-y) \mid x = -1,0,1; y = -1,0,1\} \quad (1)$$

$$S_\times(i,j) = \{(i-x, j-y) \mid x = -1,1; y = -1,1\} \quad (2)$$

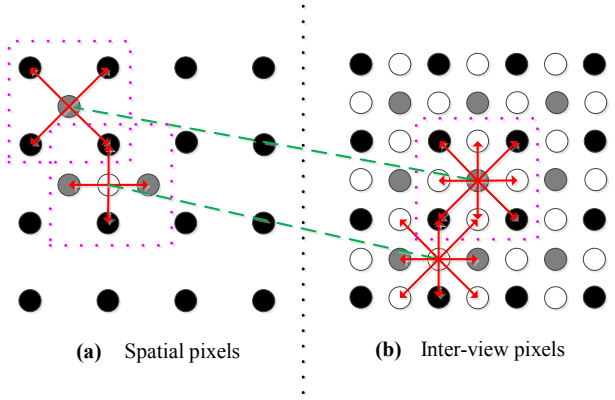
$$S_+(i,j) = \{(i, j-1), (i, j+1), (i-1, j), (i+1, j)\} \quad (3)$$

Therefore, we can express the value $p_n(i,j)$ of missing pixels as

$$p_n(i,j) = \sum_{(u,v)\in S_n} w(u,v) * x(u,v) + \sum_{(u,v)\in S_k} w(u,v) * x'(u,v) \quad (4)$$

Where x' is the pixels from the neighboring views, S_n is the set of the spatial neighbors, which is equal to $S_\times$ and S_+ in the first and second pass, respectively; S_k is the pixel set from the neighboring views; $w(u,v)$ is the weight coefficient of every reference pixel in S_n and S_k.

As the autoregressive model can adjust the weight coefficients for different areas, the spatial and inter-view correlation are taken into account simultaneously so as to make the super-resolved image more accurate and smooth. It can be seen that the weight coefficient estimation process plays a crucial role in the interpolated process. The details are described in the following subsections.



(a) Spatial pixels **(b)** Inter-view pixels

Fig.3 Formation of reference pixels in spatial and inter-view images

B. Weight Coefficient Estimation

The multi-view images captured at the same time instance shows the same objects from different viewpoints, so the objects appearing in the neighboring views also appear in the current view but at different pixel locations. According to the stationary property of the image, we assume that the weight coefficients remain the same in stereo images for some regions. To find the corresponding regions in stereo images, a block based scene-matching algorithm is employed. More specifically, for a block B in LR image, we firstly find the matching block B' in the HR image in the sense of the least square error (MSE). Because the HR image has a higher resolution than the LR image, the scene-matching process should be performed at sub-pixel precision.

For every pixel in the stereo images, we assume that auto-regressive image model is satisfied, which is expressed as follows.

$$X(i,j) = \sum_{(u,v)\in S_n} w(u,v) * X(i+u, j+v) + \sum_{(u,v)\in S_n^{'}} w(u,v) * X^{'}(i+u, j+v) + \delta(i,j) \quad (5)$$

Where $X(i,j)$ is the pixel value at position (i,j); $X^{'}(i,j)$ are the corresponding pixels from the neighboring view; $w(u,v)$ is the weight coefficient associated with the corresponding reference pixel and $\delta(i,j)$ is the random noise independent of the position (i,j). For simplicity, we represent the spatial neighboring pixel set X and the interview neighboring pixel set $X^{'}$ as B, which is $B=X \cup X^{'}$.

Therefore, based on the auto-regressive model, the weight coefficients can be estimated from the following linear least square problem.

$$W = \arg\min (X(i,j) - \sum_{(u,v)\in T} w(u,v) * B(u,v))^2 \quad (6)$$

The closed-form solution to the problem is shown as follows.

$$W = (C^{T}C)^{-1}C^{T}X \quad (7)$$

Where each row of the matrix C contains the vectors of B, which is $S_k \cup S_\times$ in the first pass and $S_k \cup S_+$ in the second pass; X is the vector consisting of the pixels in the local window T.

IV. EXPERIMENTS

To demonstrate the efficiency of the proposed method, it is tested on a series of multi-view images, which are consisted of real and synthetic images. For synthetic stereo images, we used the following images as our tests: *Cones*, *Teddy* and *Saw tooth* [16], where the right view is lower resolution and the left view is kept as full resolution. For real stereo images, we used frame 0 of the following multi-view sequences as the tests: *Pantomime* (512x384) and *Balloon* (512x384) [17]. For *Pantomime*, the view 37 is used as lower resolution view and view 39 is used as full resolution. For *Balloon*, the view 1 is chosen as lower resolution view and view 3 is used as full resolution view. To get the lower resolution images, we down-sample the high resolution version by a factor of two in both row and column dimensions through direct down-sampling method, in which the down-sampled value is the upper left one of the four corresponding intensity values.

In table I, we present the PSNR for Luma component of two versions of an up-sampled version interpolated by NEDI [13], which is an auto-regressive super-resolution method without using inter-view correlation; and an interpolated version by using the proposed method.

Table 1. PSNR of interpolation version using NEDI and the proposed method [dB]

Sequence	NEDI [13]	Proposed	ΔPSNR
Cones	28.84	29.84	1.00
Teddy	30.19	31.04	0.85
Saw tooth	28.61	30.27	1.66
Pantomime	34.58	38.88	4.30
Balloon	33.74	34.30	0.56

The proposed method achieves a significant gain in PSNR for each test image, especially for *Pantomime*. Such gains are resulted from the high frequency information from the full resolution image in the neighboring view. The details of the original, interpolated by NEDI and up-sampled by the proposed method of *Pantomime* sequence are illustrated in Fig 4, from which a great visual quality improvement can be seen on the clown's face, hat and clothes.

(a) (b) (c)

Fig.4 Details of the results of frame 0 in Pantomime view 37. (a) Original (b) NEDI (c) Proposed

Similar visual quality improvement can be achieved for synthetic stereo images. The visual quality comparison between the interpolated version by NEDI and the proposed method for *Cones* is shown in Fig 5. Great visual quality improvement can be seen on the words in the image.

(a)

(b) (c)

Fig .5 Visual quality improvement of *Cones* (a) Original (b)NEDI (c)Proposed

V. CONCLUSION

The paper proposes a super-resolution method based on auto-regressive model for interpolating lower resolution image in the mixed resolution multi-view framework. The auto-regressive model is used to exploit the inter-view and spatial correlation simultaneously. Instead of using the depth map or disparity estimate process, a block based scene matching algorithm is adopted to establish the correspondence between images in different views. The efficacy in terms of both PSNR and visual quality of the proposed method is demonstrated by the experimental results.

Future work includes the extensions of the proposed method into the existing multi-view coding standard. In this situation, the influence of the noise created during compression should be considered. In addition, the block based scene matching algorithm should be investigated further, for the accuracy of correspondence makes a great impact on the performance.

ACKNOWLEDGEMENTS

This work was supported in part by National Basic Research Program of China (973 Program 2009CB320905), in part by the National Science Foundation of China under Grants 61272386 and in part by National High Technology Research and Development Program(863), 2012AA011505.

REFERENCES

[1] Y. Chen, M. Hannuksela, L. Zhu, A. Hallapuro, M. Gabbouj, H. Li, "Coding techniques in multiview video coding and joint multiview video model," Picture Coding Symposium, pp.1-4, 6-8 May 2009.

[2] B. Julesz, Foundations of cyclopean perception, University of Chicago Press, 1971.

[3] H. Brust, A. Smolic, K. Mueller, G. Tech, T.Wiegand, "Mixed resolution coding of stereoscopic video for mobile devices," 3DTV Conference, pp.1-4, 2009.

[4] C. Fehn, P. Kauff, S. Cho, N. Hur, J. Kim, "Asymmetric coding of stereoscopic video for transmission over T–DMB," in Proc.3DTV-CON, May 2007, pp. 1–4.

[5] A. Aksay, C. Bilen, E. Kurutepe, T. Ozcelebi, G. Akar, R. Civanlar, A. Tekalp, "Temporal and spatial scaling for stereoscopic video compression," 14th European Signal Processing Conference, 2006.

[6] Y. Chen, Y. Wang, M. Gabbouj, M. Hannuksela, "Regionally adaptive filtering for asymmetric stereoscopic video coding", in Proc.IEEE Int. Symp. Circuits Syst., May 2009, pp. 2585–2588.

[7] H. Sawhney, Y. Guo, K. Hanna, R. Kumar, S. Adkins, S. Zhou, "Hybrid stereo camera: an IBR approach for synthesis of very high resolution stereoscopic image sequences," ACM SIGGRAPH, pp. 451–460, 2001.

[8] E. Ekmekcioglu, S.Worrall, A. Kondoz, "Utilisation of downsampling for arbitrary views in multi–view video coding," IEEE Electronics Letters, v. 44, pp. 339-340, 2008.

[9] Hailong Yang, Mei Yu, Gangyi Jiang, "Decoding and up-sampling optimization for asymmetric coding of mobile 3DTV", in: Proceedings of the IEEE TENCON, 2009, pp. 1–4.

[10] L. Zhang, X. Wu, "An edge-guided image interpolation algorithm via directional filtering and data fusion", IEEE Transactions on Image Processing 15 (2006) 2226–2238.

[11] W. Dong, L. Zhang, G. Shi, X. Wu, "Nonlocal back-projection for adaptive image enlargement", in: IEEE International Conference on Image Processing, Hongkong, 2010, pp. 349–352.

[12] H.-Y. Chen, J.-J. Leou, "Saliency-directed image interpolation using particle swarm optimization", Signal Processing 90 (2010) 1676–1692

[13] X. Li, M. Orchard, "New edge-directed interpolation", IEEE Transactions on Image Processing 10 (2001) 1521–1527.

[14] D.C. Garcia, C. Dorea, R.L. de Queiroz, Super-resolution for multi-view images using depth information, in: Proceedings of the IEEE International Conference on Image Processing, Hong Kong, 2010, pp. 1793–1796

[15] Jing Tian, Li Chen, Zhenyu Liu, "Dual regularization-based image resolution enhancement for asymmetric stereoscopic images", Signal Processing 92 (2012) 490–497

[16] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," Int. J. Comput. Vision,vol. 47, nos. 1–3, pp. 7–42, Apr.–Jun. 2002

[17] Nagoya University. FTV Test Sequences [Online]. Available: http://www.tanimoto.nuee.nagoya-u.ac.jp