KSII TRANSACTIONS ON INTERNET AND INFORMATION SYSTEMS VOL. 8, NO. 9, Sep. 2014 Copyright  $\odot$  2014 KSII

# Auto-Covariance Analysis for Depth Map Coding

Lei Liu, Yao Zhao, Chunyu Lin, and Huihui Bai

Institute of Information Science, Beijing Jiaotong University Beijing, China 100044 [e-mail: {12112061, yzhao, cylin, hhbai}@bjtu.edu.cn] \*Corresponding author: Yao Zhao

Received June 12, 2014; accepted July 23, 2014; published September 30, 2014

## Abstract

Efficient depth map coding is very crucial to the multi-view plus depth (MVD) format of 3-D video representation, as the quality of the synthesized virtual views highly depends on the accuracy of the depth map. Depth map contains smooth area within an object but distinct boundary, and these boundary areas affect the visual quality of synthesized views significantly. In this paper, we characterize the depth map by an auto-covariance analysis to show the locally anisotropic features of depth map. According to the characterization analysis, we propose an efficient depth map coding scheme, in which the directional discrete cosine transforms (DDCT) is adopted to substitute the conventional 2-D DCT to preserve the boundary information and thereby increase the quality of synthesized view. Experimental results show that the proposed scheme achieves better performance than that of conventional DCT with respect to the bitrate savings and rendering quality.

**Keywords:** Multi-view plus depth (MVD), depth coding, auto-covariance model, directional discrete cosine transform (DDCT), 3-D video

http://dx.doi.org/10.3837/tiis.2014.09.011

Part of this paper is submitted to the International Conference on Embedded Systems and Intelligent Technology, Sep. 18-20, 2014, Gwangju, Korea. This work was supported in part by National Basic Research Program of China (No. 2012CB316400), National Natural Science Foundation of China (No. 61025013, No. 61210006 and No. 61272051), Program for Changjiang Scholars and Innovative Research Team in University (No. IRT201206), Beijing Higher Education Young Elite Teacher Project (YETP0543), and SRFDP (20130009120038).

# 1. Introduction

**R**ecent years witnessed the significant attention of three-dimensional (3-D) videos. 3-D videos enable the viewers to select an interactive viewpoint and perceive immersive experience of the reality scenes [1]. However, the applications of 3-D video suffer from the huge amount of multi-view video data to be compressed and transmitted. To solve this problem, multi-view video coding (MVC) [2] has been proposed and standardized as an extension of the H.264/MPEG-4 AVC by the joint video team of the ITU-T video coding experts group (VCEG) and ISO/IEC moving picture experts group (MPEG). The MVC tries to exploit the inter-view correlation between different video sequences and the spatial/temporal correlation in each single sequence. Though MVC achieves tolerable compression, it still seems to be too many video data volumes to process. Another feasible solution is using the data format of multi-view video plus depth (MVD) representation [3], in which the intermediate virtual (novel) views can be generated by the transmitted video views and their corresponding depth maps. In MVD coding structure, video views along with their depth maps are encoded and transmitted; at the decoder side, a number of desired intermediate views can be synthesized from the neighboring viewpoints, via some depth-image-based rendering (DIBR) techniques [4].



Fig. 1. An example of texture image and depth image (*Cones*). (a) texture image; (b) depth image.

Depth map represents the distance information between the capturing camera and the objects in the scene. It can be considered as a gray scale image. **Fig. 1** presents a texture (color) image *Cones* and its corresponding depth image. As can be seen, the depth image contains nearly no texture but sharp object boundaries, as the gray levels are almost the same in most regions within an object but change abruptly when across the boundaries [5]. Depth map plays an important role in the virtual view synthesis. In the view synthesis, the distortion of depth data, especially around the object boundaries, will lead to geometry changes and occlusion variations of the texture image. This will seriously degrade the quality of the synthesized views [6]. Therefore, efficient depth map coding that preserves the depth information (especially the boundary fidelity) is an essential part to the 3DV systems.

In the image and video compression techniques, the two-dimensional discrete cosine transform (2-D DCT) is the most widely used transform because of its efficient energy compaction capability and computational simplicity. However, the 2-D DCT has already been proven that it is inefficient to encode image blocks with complex textures or edges [7]. Since the quality of synthesized views is very sensitive to the boundary accuracy of the depth, it seems that DCT is also not the best choice for the depth map coding.

In this paper, we analyze the local anisotropic features of depth map by an adaptive auto-covariance characterization which shows some special statistical characteristics of depth map. According to the characterization analysis, we propose an efficient depth map coding scheme using directional discrete cosine transforms (DDCT) [7] adapted to these locally anisotropic features, which can well preserve the boundary accuracy of depth maps and consequently increase the quality of the synthesized views.

The rest of the paper is organized as follows. Section 2 gives a brief overview of some related works. Section 3 firstly presents the auto-covariance analysis for the depth map and then the proposed coding scheme. Experimental results and performance comparisons are shown in Section 4, and finally Section 5 concludes the paper.

## 2. Related Works

## 2.1 Depth Map Coding

A direct approach to compress the depth map is to treat it as an ordinary image or video and process it using the existing coding standards such as JPEG, H.264/AVC [8], or the most recently emerging high efficiency video coding (HEVC) standard [9]. However, depth maps have unique characteristics different from texture/color images, which make these compression tools not suitable for depth map coding. First, depth map solely represents the distance between camera and objects, so depth levels within an object are nearly the same. It contains no texture, smooth regions within objects and the background, and discontinuous boundaries. Second, the temporal consistency of depth map is much lower than that of the color videos, because of low resolution depth capture devices or the inaccurate depth estimation method. Moreover, the depth map is never displayed; it only assists the decoder in synthesizing the virtual views. Thus, to achieve optimal synthesis results, the affection of depth distortion to the synthesized views needs careful investigation during depth map coding process [10].

Considering these characteristics, several approaches have been proposed for efficient depth map coding. [5] proposed an adaptive geometry-based intra prediction method, in which some partitioned intra prediction modes were properly produced along object boundaries to reduce the coding loss of boundary information. In [11], an edge-aware intra prediction method was produced to reduce prediction error in blocks with arbitrary edge shapes, using a graph-based representation of the pixels based on edge information. Beside these geometry-based presentation methods, [12-14] proposed some shape-adaptive transforms for efficient depth map coding. These transforms require the edge information be knowable a priori and implement along these detected edges rather than across them. These transforms produce smaller coefficients and achieve remarkable coding efficiency improvement. However, all these works do not analyze the characteristics of depth map quantitatively. Further analyses of depth map need to be investigated to quantify its characteristics.

3148

## 2.2 Directional transforms in image coding

Usually, the conventional 2-D DCT used in the image and video compression is implemented as two separable 1-D DCTs along the horizontal and vertical directions, respectively. However, there are many image blocks contain other directional information rather than the horizontal/vertical one, such as the anisotropic edges, boundaries, textures, etc. When the two 1-D DCTs implement across these edges, some unnecessary non-zero coefficients will be produced, which makes the conventional 2-D DCT not be the best choice for these image blocks [7].

Oriented information is very important to the human visual system. To achieve high coding performance, the oriented information must be exploited and preserved as much as possible. Many literatures take the directional information into account, and show significant coding gain by exploiting the directional information within images [7, 15-17]. The video coding standard H.264/AVC has developed several directional predictions (including the vertical and horizontal directions); furthermore, the most recently finalized HEVC provides up to 33 directional prediction modes for its prediction unit. Zeng and Fu proposed a directional discrete cosine transforms (DDCT) framework [7], in which the first 1-D transform of the conventional 2-D DCT is reorganized following the dominating edge direction of an image block, and the produced coefficients are rearranged appropriately to align with each other and make the second transform a horizontal one. Theoretical analysis showed that the DDCT reached a remarkable coding gain compared with the conventional DCT.

#### 3. Depth characteristic analysis and coding

In this section, we first analyze the characteristics of the depth maps. Then we briefly introduce the DDCT framework. Finally, the depth map is coded using DDCT with all the available directional modes, in which a synthesized view distortion optimization is performed to select the best DDCT mode.

#### 3.1 Auto-Covariance Analysis for Depth Map

A stationary first-order Markov signal has an auto-covariance given by:

$$R(I) = \rho^{|I|},\tag{1}$$

where I presents the distance between two elements from which the auto-covariance is computed.

Images can be modeled as Markov signals, as the value of a pixel practically depends only upon a finite number of neighboring pixels. For 2-D auto-covariance function, separable model and generalized model can be constructed from (1), which are given below in (2) and (3), respectively [18]:

$$R(I,J) = \rho_1^{|I|} \rho_2^{|J|},\tag{2}$$

$$R(\theta, I, J) = \rho_1^{|Icos(\theta) + Jsin(\theta)|} \rho_2^{|-Isin(\theta) + Jcos(\theta)|}.$$
(3)

The generalized model is a rotated case of the separable model, which can capture the local anisotropies within images. The parameter  $\theta$  represents the rotation angle.

Consider the ground truth disparity (depth) image of *Cones* shown in **Fig. 1** (b), we estimate the parameters of the two models, i.e.,  $\rho_1$  and  $\rho_2$  for the separable model, and  $\rho_1$ ,  $\rho_2$  and  $\theta$  for the generalized model. For the 8x8 size blocks of the depth map, the auto-covariance coefficient of each block is firstly estimated using the unbiased estimator. Then the parameters  $\rho_1$ ,  $\rho_2$  and  $\theta$  are found by minimizing the mean square error between the estimated auto-covariance and the models in (2) and (3). Here for intensively display of the estimated parameters,  $\rho_1$  is always chosen as the larger covariance coefficient and  $\theta$  varies between 0 and  $\pi$ . The estimation results are shown in **Fig. 2**. In the figures, each point is estimated from a block of size 8x8.



Fig. 2. Estimated auto-covariance model parameters for the depth of *Cones*. (a) separable model; (b) generalized model.

As can be seen in **Fig. 2** (a) and (b), the points in the plot from the generalized model concentrate towards the southeast region, while the points in the plot from the separable model distributes somewhat evenly. Quantitatively, most values of  $\rho_1$  in **Fig. 2** (b) are enlarged (tend to 1) while the values of  $\rho_2$  are nearly remaining the same, as compared with the values of  $\rho_1$  and  $\rho_2$  in **Fig. 2** (a). This demonstrates that the correlation of pixels oriented along an angle  $\theta$  is enhanced. This implies that the generalized model which considers the directional information within an image can provide a more faithful characterization of the image, thus better compression of the image can be expected accordingly.

## 3.2 Directional DCT

There are eight directional modes defined in the DDCT framework, following the intra prediction modes used in H.264/AVC. The directional modes are denoted as Modes 0-1 and Modes 3-8. Modes 0-1 are vertical and horizontal modes, respectively, and they come back into the conventional DCT (we denote them as Mode 0/1 here). Modes 3-8 are diagonal down-left, diagonal down-right, vertical-right, horizontal-down, vertical-left and horizontal-up, respectively. We use these notations directly in our paper. Mode 2, i.e., the planar mode, is not considered here. So there are total seven directional modes here. After the definition of the directional modes, the DDCT is conducted by performing the first 1-D

transform along the chosen direction, followed by the second 1-D transform arranged as a horizontal one. Finally, a modified zigzag scanning converts these manipulated coefficients into a 1-D sequence so as to facilitate the runlength-based VLC. Obviously, three extra bits are needed to identify the selected mode for each image block.

To show the defect of 2-D DCT and the effectiveness of DDCT, we present an example in **Fig. 3**. An artificial 8x8 block with a distinct directional (diagonal) boundary is firstly created, as shown in **Fig. 3** (a). The pixels along the boundary carry the same grey level x(i, j) = 140, while the rest pixels have the same grey level 50. Then the manual block is transformed by the 2-D DCT and DDCT (with the corresponding directional mode), and the coefficients are shown in (4) and (5), respectively:

$$c_{2D\_DCT} = \begin{bmatrix} 478.75 - 15.60 - 14.70 - 13.23 - 11.25 - 8.84 - 6.09 - 3.10 \\ -15.60 - 61.51 & 34.71 & 18.35 & 18.18 & 12.26 & 9.23 & 4.31 \\ -14.70 & 34.71 & 44.43 & -48.49 - 14.70 - 17.76 & -7.95 & -5.51 \\ -13.23 & 18.35 & -48.49 - 18.89 & 55.91 & 10.39 & 15.27 & 3.65 \\ -11.25 & 18.18 & -14.70 & 55.91 & -11.25 - 55.03 & -6.09 & -9.82 \\ -8.84 & 12.26 & -17.76 & 10.39 & -55.03 & 41.39 & 45.56 & 2.44 \\ -6.09 & 9.23 & -7.95 & 15.27 & -6.09 & 45.56 & -66.93 - 28.14 \\ -3.10 & 4.31 & -5.51 & 3.65 & -9.82 & 2.44 & -28.14 & 84.01 \end{bmatrix},$$

$$c_{DDCT} = \begin{bmatrix} 445.99 - 18.08 - 187.6651.1145.32 - 75.30 - 42.17 & 86.47 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & -12.06 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & -35.69 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 37.97 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & -70.32 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & -35.37 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & -35.37 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 82.87 \end{bmatrix}$$
(4)

The results show that the DDCT coefficients are more sparse than the 2-D DCT coefficients and also have a different energy distribution. Then the coefficients are quantized using a uniform quantizer with a quantization step (Q-Step) 30. Finally, the de-quantized values are transformed by the inverse transforms to get the reconstructed block, which are shown in **Fig. 3** (b) and (c). As can be seen in **Fig. 3** (b), the values along the boundary vary severely, and there are also many disgusting distortions around the boundary. On the other hand, **Fig. 3** (c) shows that the values along boundary are much more coincident with the original block, and with less distortion around the boundary.



**Fig. 3.** An example of the defect of 2-D DCT. (a) original block with a directional boundary; (b) reconstructed block by 2-D DCT with a quantization step 30 (MSE=95.69); (c) reconstructed block by DDCT with the same quantization step (MSE=27.43).

#### **3.3 Directional Mode Selection**

According to the analysis of Sec. 3.1, taking the directional information into account may lead to better coding performance. We use the DDCT (including 2-D DCT as a special case) for the depth map coding. Each block of the depth map is encoded with the available transform. Then a rate-distortion optimization (RDO) function is formed as in (6) to select the transform mode:

$$J_{mode}^{i} = D_{mode}^{i} + \lambda R_{mode}^{i}, \tag{6}$$

where  $D_{mode}^{i}$  is the distortion (indicated by MSE) of the *i*-th block for the current DDCT mode and  $R_{mode}^{i}$  is the amount of bits needed to encode that block, respectively. All the available modes are conducted for the block, and the one with minimum R-D cost is selected as the best mode for that block.

However, the cost function in (6) does not consider the synthesis distortion during the DIBR process and will not get the optimal performance. A newly defined synthesized view distortion optimization (SVDO) function considering both the depth map error and the virtually rendered view quality is proposed as follows [19]:

$$D_{avg}(k) \approx \frac{1}{2} \cdot \alpha \cdot \Delta D_k \cdot \left[ \left| \tilde{C}_k - \tilde{C}_{k-1} \right| + \left| \tilde{C}_k - \tilde{C}_{k+1} \right| \right],\tag{7}$$

where  $\Delta D_k$  is the depth distortion at position k,  $\tilde{C}_k$ ,  $\tilde{C}_{k-1}$  and  $\tilde{C}_{k+1}$  are color pixel values at position k, k-1 and k+1, respectively; and  $\alpha$  is a coefficient determined by the camera parameters through the following equation:

$$\alpha = \frac{f \cdot L}{255} \cdot \left(\frac{1}{Z_{near}} - \frac{1}{Z_{far}}\right),\tag{8}$$

where f is the focal length, L is the baseline between the current and rendered view,  $Z_{near}$  and  $Z_{far}$  are the nearest and farthest depth values of the scene, respectively.

This new cost function combines both the position shift by depth errors and difference between adjacent texture image pixels during the warping/rendering process. It could be give a more reasonable mode selection in terms of rendering view quality.

## **4. Experimental Results**

To validate the performance of the proposed scheme, we compare it with the coding scheme based on the conventional 2-D DCT. Five ground truth disparity images provided by Middlebury stereo datasets [20]: *Barn1*, *Cones*, *Poster*, *Sawtooth* and *Venus*, which contain piecewise planar scenes with distinct visible directional boundaries, are tested in the experiments. Among the test sets, each disparity image (depth) of the second view is selected to process, and the virtual view is warped from the second view and the processed depth. To estimate the quality of the warped synthesized view, we use the warped virtual view from the original second view and its original depth as the reference.



Fig. 4. R-D performance for the synthesized views (Bitrate-PSNR).

**Fig. 4** shows the R-D performance (PSNR vs. Bitrate) for the synthesized virtual views of the test images. As can be seen, compared with the 2-D DCT, DDCT (with both mode selection methods) achieves better compression performance at all the encoding bitrates for all the test sets. On the other hand, mode selection using the SVDO cost function outperforms that using the RDO cost function, as the former takes the rendering view quality into account.

For immediate observation, the Bjonteggard Delta PSNR (BD-PSNR) and Bjonteggard Delta Bitrate (BD-Bitrate) [21] of the synthesized views for the test depth maps between the proposed scheme and the 2-D DCT based scheme are shown in Fig. 5. The BD-PSNR measures the average vertical distance between two R-D curves [Fig. 4 (a)-(e)] and it implies

the average coding gains in terms of dB [Fig. 5 (a)]. The BD-Bitrate measures the average horizontal distance between two R-D curves [Fig. 4 (a)-(e)] and it implies the average bitrate savings in terms of percentage [Fig. 5 (b)]. These results show that the proposed scheme achieves improvement with maximum coding gain of 2.81 dB and 1.80 dB on average, or 18.40% and 13.20% of bitrate saving, respectively.



Fig. 5. The average coding gains and average bitrate savings for the synthesized views. (a) average coding gains; (b) average bitrate savings.



**Fig. 6.** The difference between the synthesized views using various compressed depths and the reference view of *Barn1*. (a) reference view synthesized using the original depth; (b) view synthesized using depth compressed by 2-D DCT; (c) view synthesized using depth compressed by DDCT+RDO; (d) view synthesized using depth compressed by DDCT+SVDO. The figures in the bottom row are the enlarged portions of the red rectangle areas in the top row.

**Fig. 6** shows the difference between the synthesized view and the reference view of *Barn1* (with an enlarged portion). It can be seen that the depth map coded by our proposed scheme produces less distortions in the synthesized view, especially in the area lightened by the red rectangle.

**Fig. 7** shows the mode selection result using the two cost functions for *Barn1*. As can be seen, the most popular scenario is Mode 0/1, i.e., the conventional 2-D DCT is selected in nearly all the smooth areas; however, around the boundaries, various directional modes are selected. This is quite corresponding to the characteristics of depth map. The directional modes distributed around the boundaries can explain the coding gains (or the bitrate savings) when using DDCT to some extent. Though the amount of directional modes is quite small compared with Mode 0/1, there are always some remarkable coding gains achieved by the directional transforms. This indicates that the fidelity of the boundary area is conclusively important to the quality of synthesized views. Moreover, as shown in **Fig. 8**, there is some difference between the two scenarios of mode distribution which use different cost functions. This is the reason why the performance of DDCT plus SVDO always outperforms that of DDCT plus RDO, as shown in **Fig. 5** (DDCT+SVDO vs. DDCT+RDO).



Fig. 7. DDCT mode distributions for *Barn1* using two cost functions. (a) RDO; (b) SVDO.



Fig. 8. The difference between the mode selection results using RDO and SVDO.

## 5. Conclusion

In this paper, the statistical characteristics of the depth map are firstly analyzed by an auto-covariance model. In order to better preserve the boundary information of the depth, an efficient depth map coding scheme is proposed using the directional discrete cosine transforms (DDCT), in which the conventional DCT is manipulated to implement along the image boundaries. A rate-distortion optimization cost function, which considers both depth errors and view synthesis errors, is adopted to select the best directional mode for each image block.

By exploiting the directional information of the boundary, experimental results show that the proposed scheme achieves significant performance improvement for depth map coding, and thereby improves the quality of the synthesized views. Future works will be focused on the analysis of influence of depth map distortion on synthesized view. A more sophisticated rate-distortion optimization metric based on the view synthesis distortion is also needed careful investigation.

#### References

- A. Smolic, K. Mueller, P. Merkle, C. Fehn, P. Kauf, P. Eisert, and T. Wiegand, "3D video and free view-point video-technologies, applications and MPEG standard," in *Proc. of IEEE Int. Conf. Multimedia and Expo* (ICME 2006), pp. 2161–2164, July 2006. <u>Article (CrossRef Link)</u>
- [2] A. Vetro, S. Yea, M. Zwicker, W. Matusik, and H. Pfister, "Overview of multiview video coding and anti-aliasing for 3D displays," in *Proc. of IEEE Int. Conf. Image Process.* (ICIP 2007), pp. I-17–I-20, Sep. 2007.<u>Article (CrossRef Link)</u>
- [3] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Multi-view video plus depth representation and coding," in *Proc. of IEEE Int. Conf. Image Process.* (ICIP 2007), pp. I-201–I-204, Sep. 2007. <u>Article (CrossRef Link)</u>
- [4] C. Fehn, "A 3D-TV approach using depth-image-based rendering (DIBR)," in Proc. of Visual., Imag., Image Process., pp. 482–487, Sep. 2003.
- [5] M.-K. Kang and Y.-S. Ho, "Depth video coding using adaptive geometry based intra prediction for 3-D video systems," *IEEE Trans. Multimedia*, Vol. 14, No. 1, pp. 121–128, Feb. 2012. <u>Article (CrossRef Link)</u>
- [6] Y. Zhao, C. Zhu, Z. Chen, and L. Yu, "Depth no-synthesis-error model for view synthesis in 3-D video," *IEEE Trans. Image Process.*, Vol. 20. No. 8, pp. 2221–2228, Aug. 2011. <u>Article (CrossRef Link)</u>
- [7] B. Zeng and J.-J. Fu, "Directional discrete cosine transforms A new framework for image coding," *IEEE Trans. Circ. Syst. for Video Technology*, Vol. 18, No. 3, pp. 305–313, Mar. 2008. <u>Article (CrossRef Link)</u>
- [8] ITU-T Rec. H.264 | ISO/IEC 14496-10 (MPEG-4 AVC), "Advanced video coding for generic audiovisual services," Mar. 2005.
- [9] G. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circ. Syst. for Video Technology*, Vol. 22, No. 12, pp. 1649–1668, Dec. 2012. <u>Article (CrossRef Link)</u>
- [10] W.-S. Kim, A. Ortega, P. Lai, D. Tian, and C. Gomila, "Depth map distortion analysis for view rendering and depth coding," in *Proc. of IEEE Int. Conf. Image Process.* (ICIP 2009), pp. 721–724, Nov. 2007. <u>Article (CrossRef Link)</u>
- [11] G. Shen, W.-S. Kim, A. Ortega, J. Lee, and H. Wey, "Edge-aware intra prediction for depth-map coding," in *Proc. of IEEE Int. Conf. Image Process.* (ICIP 2010), pp. 3393–3396, Sep. 2010. <u>Article (CrossRef Link)</u>
- [12] M. Maitre and M. N. Do, "Joint encoding of depth image based representation using shape-adaptive wavelets," in *Proc. of IEEE Int. Conf. Image Process.* (ICIP 2008), pp. 1768–1771, Oct. 2008. <u>Article (CrossRef Link)</u>
- [13] G. Shen, W.-S. Kim, S. K. Narang, A. Ortega, J. Lee, and H. Wey, "Edge-adaptive transforms for efficient depth map coding," in *Proc. Picture Coding Symp.* (PCS 2010), pp. 566–569, Dec. 2010. <u>Article (CrossRef Link)</u>
- [14] W.-S. Kim, S. K. Narang, and A. Ortega, "Graph based transforms for depth video coding," in Proc. of Int. Conf. Acoustics, Speech and Signal Process. (ICASSP 2012), pp. 813–816, Mar. 2012. Article (CrossRef Link)
- [15] J. Xu, B. Zeng, and F. Wu, "An overview of directional transforms in image coding," in *Proc. of IEEE Int. Symp. Circuits and Systems* (ISCAS 2010), pp. 3036–3039, May 2010. <u>Article (CrossRef Link)</u>

- [16] Z. Gu, W. Lin, B.-S. Lee, and C. T. Lau, "Rotated orthogonal transform (ROT) for motion-compensation residual coding," *IEEE Trans. Image Process.*, Vol. 21, No. 12, pp. 4770–4781, Dec. 2012. <u>Article (CrossRef Link)</u>
- [17] L. Liu, A. Wang, K. Zhu, C. Lin, and Y. Zhao, "Directional block compressed sensing for image coding," in *Proc. of IEEE Int. Conf. Circ. Syst.* (ISCAS 2013), pp. 1644–1647, May 2013. <u>Article (CrossRef Link)</u>
- [18] F. Kamisli and J. S. Lim, "1-D transforms for the motion compensation residual," *IEEE Trans. Image Process.*, Vol. 20, No. 4, pp. 1036–1046, April 2011. <u>Article (CrossRef Link)</u>
- [19] B. T. Oh, J. Lee, and D. Park, "Depth map coding based on synthesized view distortion function," IEEE J. of Sel. Topics in Signal Process., Vol. 5, No. 7, pp. 1344–1352, Nov. 2011. Article (CrossRef Link)
- [20] Middlebury Stereo Datasets, http://vision.middlebury.edu/stereo/data/
- [21] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," *Tech. Rep.* VCEG-M33, April 2001.



**Lei Liu** received the B.E. degree form Shandong University of Science and Technology, Qingdao, China, in 2009, and the M.E. degree form Taiyuan University of Science and Technology, Taiyuan, China, in 2012. He is currently working toward the PhD degree at Beijing Jiaotong University, Beijing, China. His current research interests include image/video compression and processing.



Yao Zhao received the BS degree from Fuzhou University, China, in 1989, and the ME degree from Southeast University, Nanjing, China, in 1992, both from the Radio Engineering Department, and the PhD degree from the Institute of Information Science, Beijing Jiaotong University (BJTU), China, in 1996. He became an associate professor at BJTU in 1998 and became a professor in 2001. From 2001 to 2002, he was a senior research fellow with the Information and Communication Theory Group, Faculty of Information Technology and Systems, Delft University of Technology, Delft, The Netherlands. He is currently the director of the Institute of Information Science, BJTU. His current research interests include image/video coding, digital watermarking and forensics, and video analysis and understanding. He serves on the editorial boards of several international journals, including as associate editors of IEEE Transactions on Cybernetics, IEEE Signal Processing Letters, and an area editor of Signal Processing: Image Communication (Elsevier), etc. He was named a distinguished young scholar by the National Science Foundation of China in 2010, and was elected as a Chang Jiang Scholar of Ministry of Education of China in 2013. He is a senior member of the IEEE and a fellow of IET.



**Chunyu Lin** was born in Liaoning Province, China. He works as a lecturer in Beijing Jiaotong University. He obtained his doctor degree in Beijing Jiaotong University in 2011. From 2009 to 2010, he was a visiting researcher at the ICT group of Delft University of Technology, Netherlands. From 2011 to 2012, He was a postdoc in Gent University, Belgium. His research interests are in the areas of image/video compression and robust transmission, stereo matching and 3D video coding.



**Huihui Bai** received the Ph.D. degrees in signal and information processing from Beijing Jiaotong University (BJTU), Beijing, China, in 2008. She is currently an associate professor in Institute of Information Science in BJTU. She has been engaged in R&D work in video coding technologies and standards such as HEVC, 3D video compression, multiple description video coding (MDC) and distributed video coding (DVC). She is leading or participating in several research projects such as 973 Program, 863 Program, National Natural Science Foundation of China, Beijing Natural Science Foundation and Jiangsu Provincial Natural Science Foundation.