

# 3D View Synthesis with Feature-Based Warping

**Ningning Hu, Yao Zhao, Huihui Bai**

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

*Received January 5, 2017; revised April 27, 2017; accepted June 1, 2017;  
published November 30, 2017*

---

## **Abstract**

Three-dimensional video (3DV), as the new generation of video format standard, can provide the viewers with a vivid screen sense and a realistic stereo impression. Meanwhile the view synthesis has become an important issue for 3DV application. Differently from the conventional methods based on depth, in this paper we propose a new view synthesis algorithm, which can employ the correlation among views and warp in the image domain only. There are mainly two contributions. One is the incorporation of sobel edge points into feature extraction and matching, which can obtain a better stable homography and then a visual comfortable synthesis view compared to SIFT points only. The other is a novel image blending method proposed to obtain a better synthesis image. Experimental results demonstrate that the proposed method can improve the synthesis quality both in subjectivity and objectivity.

---

**Keywords:** View synthesis, Image warping, Image blending, 3D video

## 1. Introduction

Three-dimensional video (3DV) equipped with stereo and interactive function has gradually achieved the viewer's favour for the vivid scene representation. As an important data format of 3DV, Multi-view Video (MVV) is generated through the videos captured for the same scene from cameras located in the different positions. In view of the huge data amount, it needs large storage or bandwidth to store or transmit the whole data of MVV. As a result, the method 3D view synthesis is demanded, by which some views can be synthesized by their adjacent views.

One of the current state-of-art view synthesis algorithms is Depth-image-based rendering (DIBR) [1], which can generate arbitrary view synthesis image with high efficiency. MPEG has formulated corresponding standardization for DIBR to handle Multi-view Video plus Depth (MVD) [2] supporting Free Viewpoint Television (FTV). Texture images with corresponding depth images and the indispensable camera parameters are required to convert  $M$  to  $N$  views ( $M < N$ ). Meanwhile, the estimation for camera parameters is a time-consuming process if there is no data available. Furthermore, the accuracy of the estimated depth images is an ill-posed task. Because of the change of occlusion relations in adjacent views, there exist so many hole regions or overlap regions [3].

On the other hand, Stefanoski et al. [4]-[5] applied image domain warping (IDW) technique to solve view synthesis task, which conducts the whole synthesis process in image domain only, without the participation of depth image. Non-linear disparity mapping [6] is the core technology applied in IDW. Though equipped with automation and high visual quality, it is a complicated process with sparse disparities to be found, saliency maps to be generated, regular grid meshes to be covered, sparse least square systems to be solved, and interpolation functions to be utilized.

As an alternative to IDW, Tao et al. [7] proposed a novel image warping algorithm which applied a global homography  $H$  as a simple model to describe the relationship of the two adjacent views. Based on the theory in [8] and [9], the homography could be illustrated as a suitable model to deal with the transformation between two different views of the same 3D scene. Compared to DIBR and IDW, this new warping method didn't rely on depth estimation or sparse linear equations calculation, but applied the robust SIFT features, which can work out an expected homography rapidly. Nevertheless, it did not consider the edge line features to conduct line matching [10-11], which are particularly important for stereo matching. And the homography had randomness to some extent during the solution of the optimum model. A bad homography may lead to a serious shape deformation. In addition, because of different fields of view in the reference image and target image, empty regions may occur on the left or right border of the synthesis result.

To address these problems, this paper proposes an improved synthesis scheme with two main contributions. The first contribution is in feature extraction. Besides the SIFT keypoints [12], sobel edge points [13] are being extracted to conduct line matching. As a good representation for the image lines, it can make the homography more stable. As the second one, a novel blending decision method is designed, which aims at dealing with the empty regions on the image border for a higher image synthesis quality.

The remainder of the paper is organized as follows. Section 2 introduces our proposed method. Both subjective and objective experimental results are shown in Section 3, and Section 4 serves as a summary for the paper.

## 2. Proposed Algorithm

The overview of the proposed view synthesis framework is shown in Fig. 1. It consists of 3 main modules: data extraction, warping calculation and image blending. In data extraction process, besides SIFT points, extra sobel edge points, which describe image line features well, are incorporated to conduct stereo matching. From Fig. 1, two kinds of matches can be found between left and middle view and between right and middle view, respectively. These matches are then used to calculate the suitable homography,  $H_1$  and  $H_2$ . Finally, a novel blending method is put forward to settle the ghost artifact caused by roughly overlaying left synthesis view with right synthesis view. Next, we will mainly explain the image warping and blending in details.

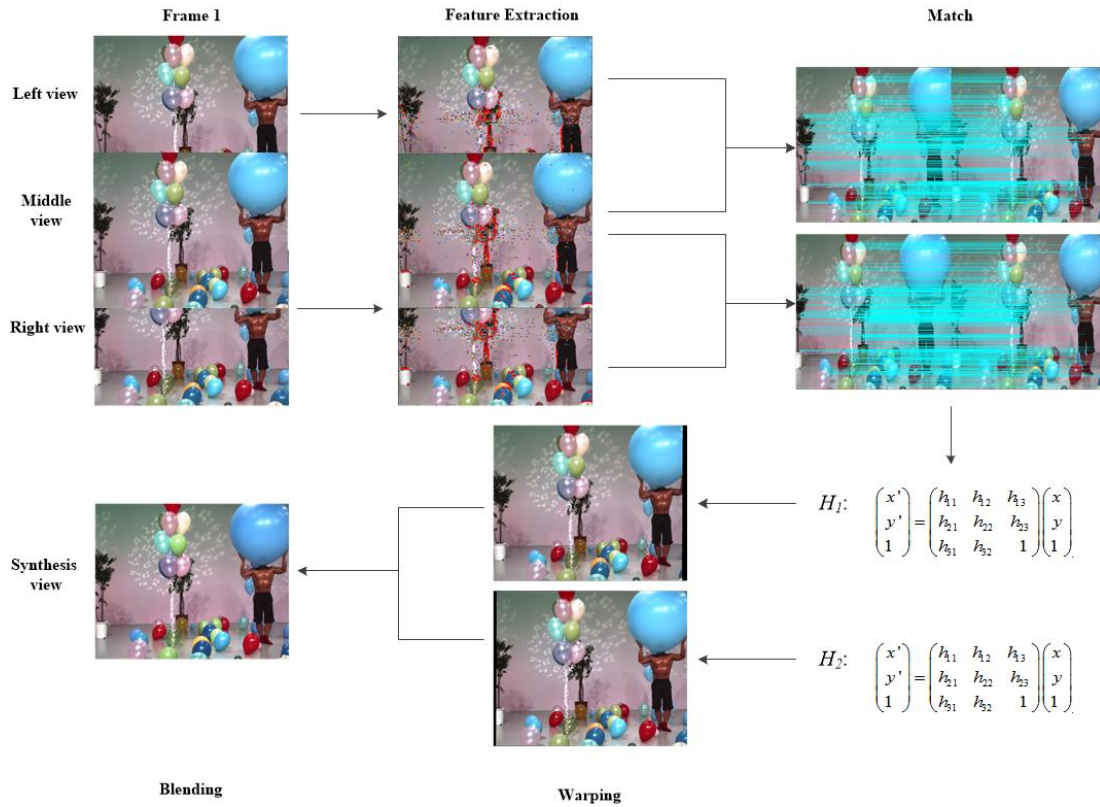


Fig. 1. Overview of the proposed featured-based 3D view synthesis framework

### 2.1 Image Warping

Image warping arises in many image analysis problems in computer vision, such as image rectification, image alignment, or image rendering. It is a transformation which maps all positions in one image plane to positions in a second image plane [14]. Homography maybe a suitable model to describe the warping relations, where the matched SIFT points are often used to get the model. In this paper, we extend the feature points with sobel edge points. Sobel edge points are resonable description for image lines, and line matching could strengthen the stability of the homography. And then the homography  $H$  obtained from the matched points is utilized to synthesize the target view.

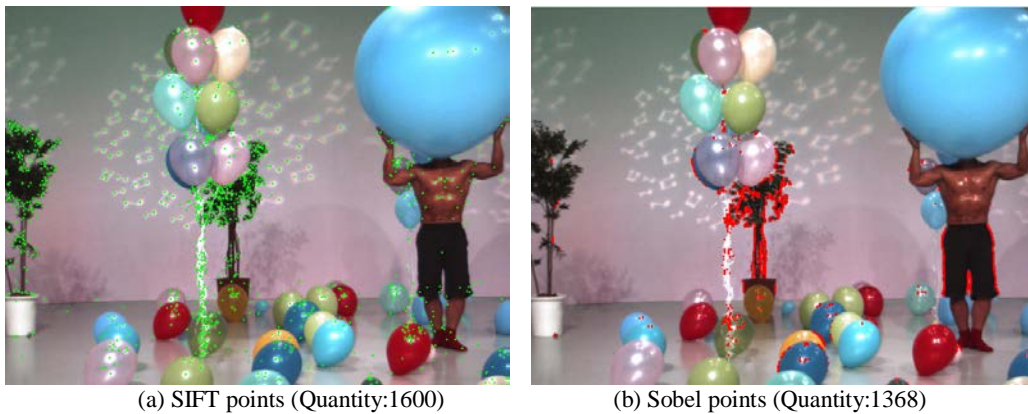
### 2.1.1 Feature Extraction

SIFT algorithm is applied extensively in the local feature representation field for its robust performance:

- **Stability.** Besides invariant to the scale, rotation and brightness, SIFT can also maintain stability for the view changing, affine transformation and noise effects.
- **Rapidity.** The SIFT algorithm after optimized can run in a rapid speed, nearly in real time.
- **Expansibility.** For every SIFT keypoint, the return value is a  $1 \times 4$  vector indicating its location and a corresponding descriptor vector of  $1 \times 128$ . Therefore, it is feasible to combine other features and constraint its descriptor vectors size as  $1 \times 128$ .

Although SIFT is characterized by the high accuracy for the keypoints, there is also a weakness derived from these points. Some keypoints is so sparse that they could not reflect the significant edge information of the object in the image [4,15].

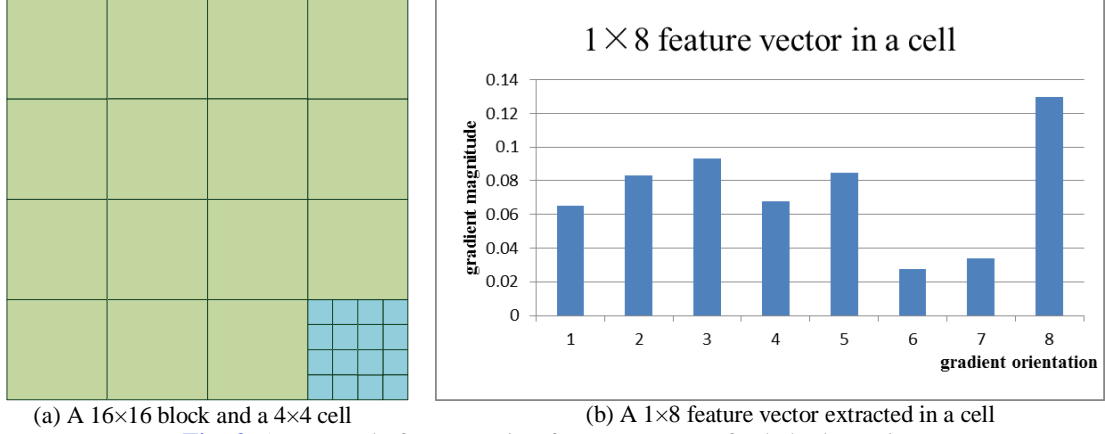
As a result, it is essential to provide a stronger constraint, which can guide a more accurate warp process. For this reason, we detected additional features distributed on the image edges. Sobel operator is used and the convolution templates, both in horizontal and vertical directions, are selected as  $3 \times 3$ . By setting the suitable gradient threshold and the step of the template window we can obtain a sobel points set which has the same order of magnitude. From Fig. 2 we can see the sobel points not only reflect the line feature that SIFT don't have (the persons black shorts) but also enhance the feature SIFT show (the pot plant behind the balloons).



**Fig. 2.** Two feature points used in data extraction

To combine with the SIFT points, every edge point is required to generate a corresponding descriptor with a size of  $1 \times 128$ . To meet this requirement, we use the model in [16]. At first a  $16 \times 16$  block centered by an edge point is extracted and the block is divided to sixteen  $4 \times 4$  cells, as clearly shown in Fig. 3(a). In the cell, every pixel point is projected to a certain column of a histogram according to its gradient orientation. And the histogram is obtained by evenly dividing  $[0, 360^\circ]$  into eight columns, we can see from Fig. 3(b) where 1-8 in abscissa represents the gradient orientation ranging from  $(0, 45^\circ]$  to  $(225, 360^\circ]$ . Then the column value can be calculated by accumulating the gradient magnitude of the pixel points within the

column. At last a  $16 \times 8 = 128$  feature vector can be gained for every edge points. We need to normalize the vector further to make it invariant to illumination.



**Fig. 3.** An example for extracting feature vectors of sobel edge points

### 2.1.2 Feature Matching

After the data extraction, SIFT keypoints combined with the sobel edge points are obtained between reference and target image views, both with location and corresponding description information.

The traditional match for each keypoint in one image is to find its nearest neighbor in the keypoints of the other image. The nearest neighbor is defined as the keypoint with the minimum Euclidean distance for the descriptor vector. The Euclidean distance is calculated as followings.

$$D = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_N - y_N)^2} \quad (1)$$

where  $X(x_1, x_2, \dots, x_N)$  and  $Y(y_1, y_2, \dots, y_N)$  is the keypoints to be matched. Meanwhile a drawback is that as long as we calculate the minimum Euclidean distance, no matter the two keypoints matched or not, there will always be a matched pair, which probably resulting in error matching.

Therefore, we use the method in [12], which can find a good match by comparing the distance of the closest neighbor to that of the second-closest neighbor. The ratio of closest to second-closest neighbors can be calculated by the following equation,

$$R = \frac{D_1}{D_2} \quad (2)$$

in which  $D_1$  represents the closet neighbor distance and  $D_2$  is the second one. Then if  $R$  is less than the ratio threshold, it means there is a low similarity between them. We regard the match with the nearest neighbor as a correct one.

### 2.1.3 Homography Calculation

Images from different views can be mapped by projection transformation. And a suitable homography matrix  $H$  can be used to fit the transform model. Suppose that the reference



image is  $R(x, y)$ , the synthesis image is  $S(x', y')$ . The relationship between  $H$  and the two images can be formulated as

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (3)$$

$H$  is a  $3 \times 3$  matrix which has 8 degrees-of-freedom, so at least 4 pairs of matching points are required to solve  $H$  matrix by linear algorithm. Because of the potential error matching points, the  $H$  calculated by the means has great randomness and instability.

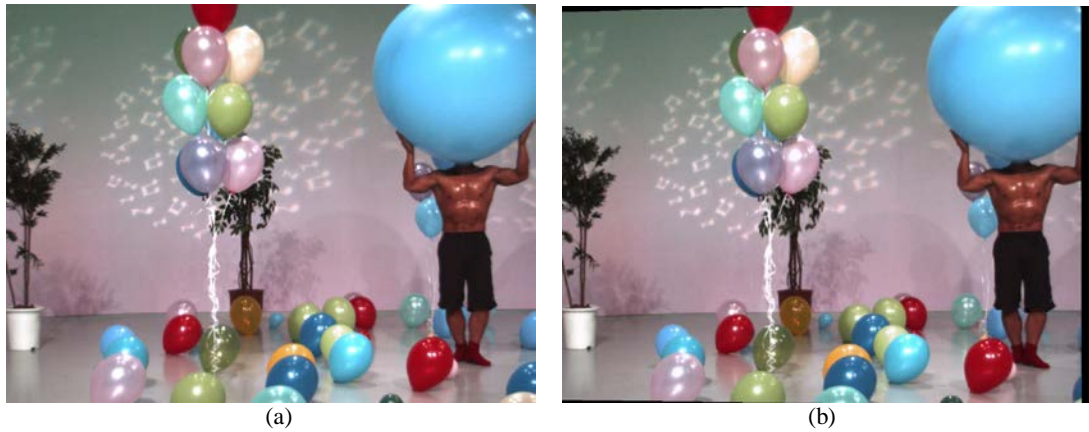
To solve the problem, random sample consensus (RANSAC) method is utilized [17]. For the RANSAC method makes full use of the data set and excludes the outer points, we can obtain a high-precision estimate model in most situations.

### 2.1.4 Warping

After we obtain a suitable homography matrix  $H$ , the synthesis image  $S(x', y')$  can be generated according to

$$S(x', y') = H * R(x, y) \quad (4)$$

**Fig. 4** shows a synthesis result with the single reference image. Though there are no discontinuous or occluded regions, the empty regions are exposed in the image border.



**Fig. 4.** An example for empty border due to depending on one-sided view.

(a) Middle view to be synthesize; (b) Synthesis result

To solve the empty regions in the image border, we introduce image blending which can use the border information of the target image to compensate the void content in the synthesis results.

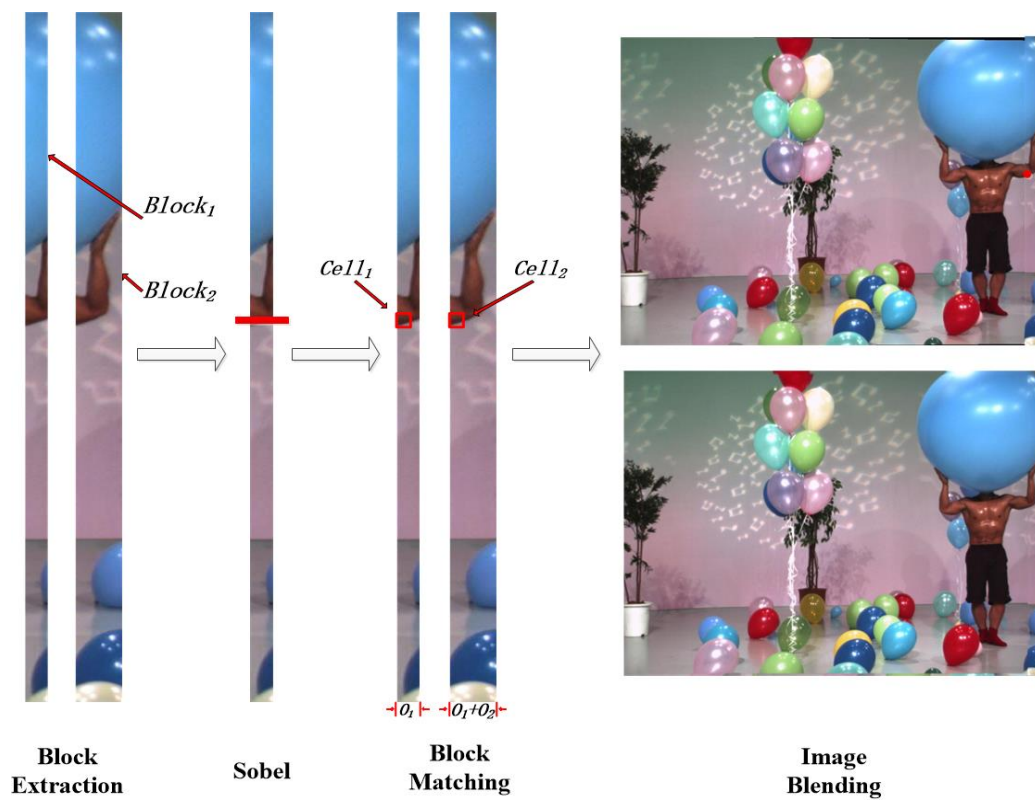
### 2.2 Image Blending

In [7], it uses only one side view to produce the middle view, which results in empty border inevitably, as **Fig. 4** shown. Many algorithms have been proposed to solve the problem and the most common solution is to use both sides to synthesize the middle view. However, it is easy to suffer ghost artifact because of inaccuracy alignment in pixel level of the two synthesis images, as **Fig. 5** shown.



**Fig. 5.** Ghost artifact by blending two views together

As the second contribution in this paper we introduce a novel blending method which can employ the border information in the target image to cover the empty content of the synthesis border. It not only fills up the empty border well, but also protects the continuity of the image content with the one-side synthesis algorithm. The proposed blending process has been shown in **Fig. 6**.



Firstly a border block  $Block_1$  with size of  $O_1 * row$  is extracted in the reference image and the same deal with the target image for a border block  $Block_2$  in a size of  $(O_1 + O_2) * row$ , both in vertical direction (horizontal direction with the same process theory).  $O_1$  is a base offset value determined by experience.  $O_2$  is the average vertical disparity offset value derived from the previous matched feature points. Theoretically speaking, we can find the corresponding content of  $Block_1$  in  $Block_2$ .

Considering that the image border always presents as small size, deficient content, it is hard to apply the traditional matching-stitching method based on feature to match the two blocks for there may be no or very few feature points to extract. To complete the matching process for the two border block, correlation matching method based on image texture is proposed. By locating the row with the maximal sobel gradient sum in  $Block_1$ , a model with  $O_1 * O_1$  around this row is build up, which we called  $Cell_1$ . We then use the correlation matching method to find out the best matching region for  $Cell_1$  in  $Block_2$ , which we called  $Cell_2$ . We traverse each pixel in  $Block_2$  from left to right, top to bottom, to derive a matching model  $Cell_2$  to be matched with  $Cell_1$ .

The correlation level  $C$  is computed as

$$C = \frac{\sum_{i=1}^n \sum_{j=1}^n Cell_1(i, j) * Cell_2(i, j)}{\left[ \sum_{i=1}^n \sum_{j=1}^n Cell_1(i, j)^2 \sum_{i=1}^n \sum_{j=1}^n Cell_2(i, j)^2 \right]^{1/2}} \quad (5)$$

In the formula,  $n$  is the height and width of the cells, which is equal to  $O_1$ ;  $Cell_1(i, j)$  and  $Cell_2(i, j)$  are the intra pixels of the cells.

Finally,  $H$  calculated before is used to calculate the new location of the model in synthesis image. According to the matching location information we can blend the one-view synthesis image with  $Block_2$ .

### 3. Experimental Results

We evaluate the performance of our novel view synthesis scheme in four standard video sequences: *Kendo*, *Balloons*, *Newspaper* and *Lovebird*, all with a resolution of  $1024 \times 768$ . In our experiments, the base offset  $O_1$  has 10 pixels, which also floats up or down with one or two pixels considering the synthesis results.

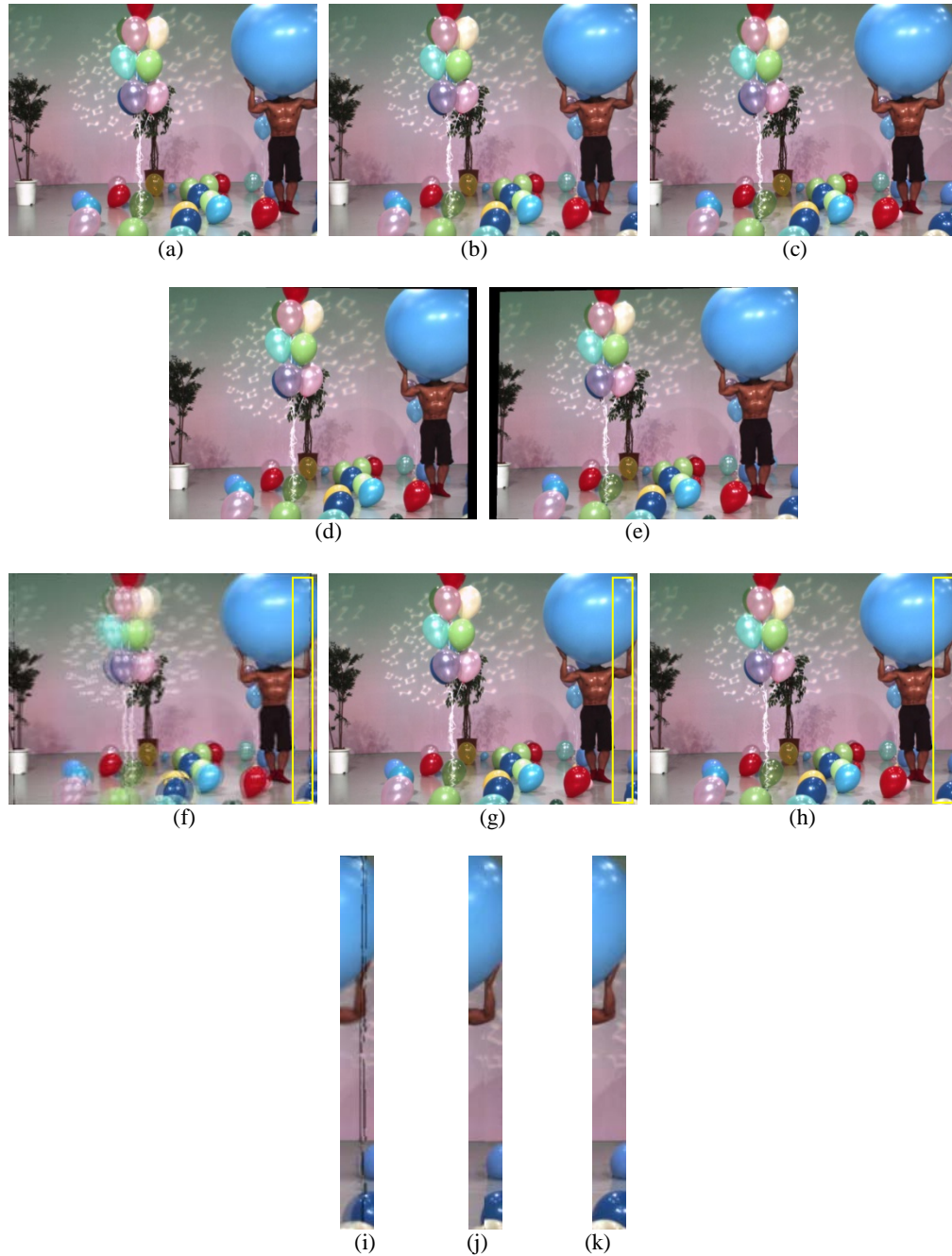
Subjectively, the synthesis effect is demonstrated in **Fig. 7-Fig. 10** along with the original images. From every third row we can see our blending algorithm compensate the empty border well by introduce block filling algorithm based on correlation matching. Objectively, we



utilizes one classical assessed standard to measure the synthesis quality, with the corresponding original image as ground truth.



**Fig. 7.** The synthesis results of *Kendo*. (a) Left view; (b) Middle view; (c) Right view; (d) Left synthesis result from [7]; (e) Right synthesis result from [7]; (f) Traditional blending; (g) Our synthesis result with left view; (h) Our synthesis result with right view; (i) Partial enlarged view of (f); (j) Partial enlarged view of (g); (k) Partial enlarged view of (h).



**Fig. 8.** The synthesis results of *Balloons*. (a) Left view; (b) Middle view; (c) Right view; (d) Left synthesis result from [7]; (e) Right synthesis result from [7]; (f) Traditional blending; (g) Our synthesis result with left view; (h) Our synthesis result with right view; (i) Partial enlarged view of (f); (j) Partial enlarged view of (g); (k) Partial enlarged view of (h).



**Fig. 9.** The synthesis results of *Newspaper*. (a) Left view; (b) Middle view; (c) Right view; (d) Left synthesis result from [7]; (e) Right synthesis result from [7]; (f) Traditional blending; (g) Our synthesis result with left view; (h) Our synthesis result with right view; (i) Partial enlarged view of (f); (j) Partial enlarged view of (g); (k) Partial enlarged view of (h).





**Fig. 10.** The synthesis results *Lovebird*. (a) Left view; (b) Middle view; (c) Right view; (d) Left synthesis result from [7]; (e) Right synthesis result from [7]; (f) Traditional blending; (g) Our synthesis result with left view; (h) Our synthesis result with right view; (i) Partial enlarged view of (f); (j) Partial enlarged view of (g); (k) Partial enlarged view of (h).

**Table 1.** The comparison between Tao's and our proposed algorithm

Test sequence	Blending Warping	Left view	Blending	$\Delta_{2L}$	Right view	Blending	$\Delta_{2R}$
Kendo	SIFT	18.7901	<b>23.6955</b>	<b>4.9054</b>	18.9005	<b>25.2782</b>	<b>6.3777</b>
	SIFT & Sobel	19.8249	<b>25.7551</b>	<b>5.9302</b>	19.6552	<b>25.4857</b>	<b>5.8305</b>
	$\Delta_1$	<b>1.0348</b>	<b>2.0596</b>	<b>6.9650</b>	<b>0.7547</b>	<b>0.2075</b>	<b>6.5852</b>
Balloons	SIFT	17.5459	<b>21.6720</b>	<b>4.1261</b>	18.0013	<b>20.9484</b>	<b>2.9471</b>
	SIFT & Sobel	18.2014	<b>22.2924</b>	<b>4.0910</b>	18.4630	<b>20.3606</b>	<b>1.8976</b>
	$\Delta_1$	<b>0.6555</b>	<b>0.6204</b>	<b>4.7465</b>	<b>0.4617</b>	<b>-0.5878</b>	<b>2.3593</b>
Newspaper	SIFT	19.4846	<b>21.3546</b>	<b>1.8700</b>	18.2448	<b>19.5113</b>	<b>1.2665</b>
	SIFT & Sobel	19.5600	<b>23.3331</b>	<b>3.7731</b>	19.8496	<b>21.4879</b>	<b>1.6383</b>
	$\Delta_1$	<b>0.0754</b>	<b>1.9785</b>	<b>3.8485</b>	<b>1.6048</b>	<b>1.9766</b>	<b>3.2431</b>
Lovebird	SIFT	25.0437	<b>26.3247</b>	<b>1.2810</b>	22.8167	<b>26.0869</b>	<b>3.2702</b>
	SIFT & Sobel	25.1842	<b>26.6742</b>	<b>1.4900</b>	23.9608	<b>26.6887</b>	<b>2.7279</b>
	$\Delta_1$	<b>0.1405</b>	<b>0.3495</b>	<b>1.6305</b>	<b>1.1441</b>	<b>0.6018</b>	<b>3.8720</b>

The objective experimental results in **Table 1**, which shows a comparable results for our proposed algorithm to [7]. The additional image quality assessed standard *PSNR* (Peak Signal to Noise Ratio) is used here. The *PSNR* is calculated as the following two formulas:

$$MSE = \frac{1}{w \times h} \sum_{i=1}^w \sum_{j=1}^h (S(i, j) - R(i, j))^2 \quad (6)$$

$$PSNR = 10 \times \lg \left( \frac{(2^n - 1)}{MSE} \right) \quad (7)$$

*MSE* (Mean Squared Error) is the mean squared error between the synthesis image  $S(x', y')$  and the reference image  $R(x, y)$ , with  $w$  and  $h$  are the width and height of the original image.  $(2^n - 1)$  represents potential biggest value for the image pixel, and  $n$  is the image's bit depth. In our experiment the image's bit depth  $n$  is 8, so the  $(2^n - 1)$  is 255.

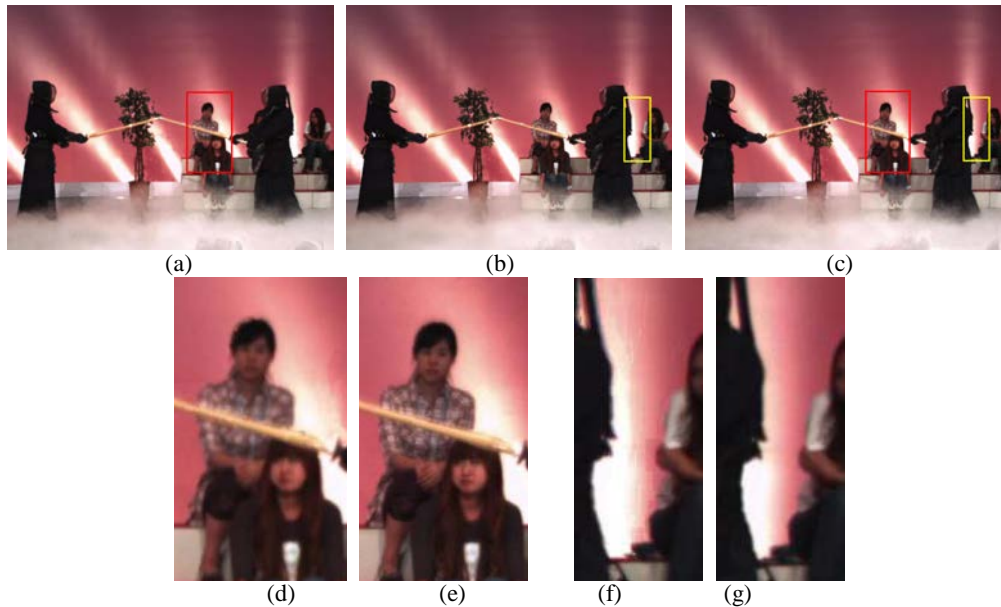
As listed in the table,  $\Delta_1$  represents a better performance after sobel edge points added.

$\Delta_{2L}$  means the *PSNR* gain by blending method with the left reference image and  $\Delta_{2R}$  is the right one. **Table 1** has shown the gains from two proposed contributions in detail. Take left view of *Kendo* as an example. From the table we can find that compared with SIFT only method [7], the incorporation of sobel as the first contribution can achieve 1.0348dB directly and 2.0596dB gains with blending method, as  $\Delta_1$  shows. Furthermore, compared with SIFT only method [7], the blending method can obtain 4.9054dB gains directly and 5.9302dB gains with the incorporation of sobel, as  $\Delta_{2L}$  shows. The two contributions can lead to 6.9650dB.



To sum up, compared to the results of [7], when we introduce sobel edge points there is a gain ranging from 0.0754dB to 1.6048dB as  $\Delta_1$  shows, when we introduce image blending there is a gain ranging from 1.2665dB to 6.3777dB as  $\Delta_{2L} (\Delta_{2R})$  shows, when the two methods are added together, there is a gain from 1.6305dB to 6.9350dB, which sufficiently demonstrate the superiority of our proposed algorithm.

In addition, to make our algorithm more convincing we also give a subjective comparison on visual level with the major view synthesis method DIBR. As Fig. 11 shows, (a) is the synthesis result from original DIBR algorithm[1]. (b) is from View Synthesis Reference Software 4.0 (VSRS 4.0) [18], which is developed by MPEG for view synthesis. Its kernel algorithm is DIBR but with some enhanced strategies. We can see although there is some improvement in (b), both can not avoid the hole filling caused by discrete pixel-level projection. Meanwhile benefitting from the continuations warping by the homography in our scheme, our results in (c) seems more smoothness.



**Fig. 11.** The subjective image quality between DIBR and ours. (a) Result from original DIBR algorithm; (b) Result from VSRS; (c) Result from our scheme; (d) and (e) are the enlarged view of (a) and (c), respectively; (f) and (g) are the enlarged view of (b) and (c), respectively.

#### 4. Conclusion and Future Work

In this paper, we propose a novel synthesis approach with feature-based warping which uses the sparse features to conduct view mapping. There are mainly two contributions in this paper. The first contribution is located in data extraction process. We combine SIFT features with sobel edge-points. The added points not only expand the amount of feature points but also make the calculation for  $H$  more robust at the expense of slight delay in time consume. As the second contribution, by using the novel blending, we work out the empty region occurred in the synthesis image by the one-sided view. Experimental results show that the synthesis views generated from our proposed method presents high quality with less ghost artifact.

In the current multi-view-coding (MVC) frameworks, texture and corresponding depth images are both necessary to synthesize the intermediate view, which is a great challenge for data storage and transmission. In the future we may attempt to realize multiview video coding by employing the homography  $H$  instead of the depth map, which means we can adapt our scheme to the current codec frame work. The basic idea is as follows. At the encoder, for the main view the standard intra coding can be applied to obtain high reconstruction quality while for its neighboring views feature descriptors can be utilized to calculate homography  $H$  between views. And then the parameters of  $H$  can be transmitted by very low bit rate and achieve the preliminary reconstruction.

## References

- [1] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," in *Proc. of SPIE, Stereoscopic Displays and Virtual Reality Systems XI*, 93, vol. 5291, May 2004. [Article \(CrossRef Link\)](#)
- [2] A. Smolic, K. Mueller, P. Merkle, N. Atzpadin, C. Fehn, M. Muller, O. Schreer, R. Tanger, P. Kauff, and T. Wiegand, "Multi-view video plus depth (MVD) format for advanced 3D video systems," *JVT of ISO/IEC MPEG & ITU-T VCEG JVT-W100*, Apr. 2007. [Article \(CrossRef Link\)](#)
- [3] T. Zarb, C.J. Debono, "Depth-based image processing for 3d video rendering applications," in *Proc. of IWSSIP 2014*, IEEE, pp.215-218, May 2014. [Article \(CrossRef Link\)](#)
- [4] N. Stefanoski, O. Wang, M. Lang, P. Greisen, S. Heinzle, and A. Smolic, "Automatic view synthesis by image-domain-warping," *IEEE Transactions on Image Processing*, vol. 22(9), pp. 3329-3341, Sep. 2013. [Article \(CrossRef Link\)](#)
- [5] M. Farre, O. Wang, M. Lang, N. Stefanoski, A. Hornung, and A. Smolic, "Automatic content creation for multi-view autostereoscopic displays using image domain warping," in *Proc. of 2011 IEEE International Conference on Multimedia and Expo*, IEEE, pp. 1–6, Jul. 2011. [Article \(CrossRef Link\)](#)
- [6] A. Smolic, Y. Wang, N. Stefanoski, M. Lang, A. Hornung, and M. Gross, "Non-linear warping and warp coding for content-adaptive prediction in advanced video coding applications," in *Proc. of 2010 IEEE International Conference on Image Processing*, IEEE, pp. 4225–4228, Sep. 2010. [Article \(CrossRef Link\)](#)
- [7] W. Tao, H. Bai, M. Liu and Y. Zhao, "Asymmetric multiview image coding based on feature matching," in *Proc. of SPIE, Optoelectronic Imaging and Multimedia Technology III*, 92732C, vol.9273, Oct.2014. [Article \(CrossRef Link\)](#)
- [8] R. Hartley, A. Zisserman, [Multiple view geometry in computer vision], Cambridge University Press, 2003. [Article \(CrossRef Link\)](#)
- [9] Z. Gang, Y. Ping, Y. He, "A new inter-view prediction method for multi-view video coding," in *Proc. of 2007 IEEE Workshop on Signal Processing System*, IEEE, pp. 227-340, Oct. 2007. [Article \(CrossRef Link\)](#)
- [10] B. Fan, F. Wu, Z. Hu, "Robust line matching through line-point invariants," *Pattern Recogn*, vol.45(2), pp. 794-805, 2012. [Article \(CrossRef Link\)](#)
- [11] Q. Jia, X. Gao, X. Fan, Z. Luo, H. Li, and Z. Chen, "Novel coplannar line-points invariants for robust line matching across views," *European Conference on Computer Vision*, Springer International Publishing, pp. 599-611, 2016. [Article \(CrossRef Link\)](#)
- [12] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol.60(2), pp. 91-110, Nov.2004. [Article \(CrossRef Link\)](#)
- [13] I. Sobel, G. Feldman, "A isotropic 3x3 gradient operator for image processing," *a talk presented at the Standard Artificial Intelligence project*, pp. 271-272, 1968. [Article \(CrossRef Link\)](#)
- [14] C.A. Glasbey, K.V. Mardia, "A review of image-warping methods," *Journal of Applied Statistics*, vol. 25(2), pp. 155-171, 1998. [Article \(CrossRef Link\)](#)

- [15] H. Wang, X. Zhang, H. Xiong, "Spatio-temporal coherence for 3-D view synthesis with curve-based disparity warping," *Visual Communications and Image Processing conference*, IEEE, pp.177-180, Dec. 2014. [Article \(CrossRef Link\)](#)
- [16] X. Xiao, X. Wang, S. Wang, "Research on image matching based on edge point feature," *International Conference on Remote Sensing*, Oct. 2010. [Article \(CrossRef Link\)](#)
- [17] M.A. Fischler, R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol.24(6), pp. 381-395, Jun.1981. [Article \(CrossRef Link\)](#)
- [18] K. Wegner, O. Stankiewicz, M. Tanimoto, M. Domanski, "Enhanced view synthesis reference software (VSRS) for free-view television," *ISO/IEC JTC1/SC29/WG11 MPEG2013/M31520*, Oct. 2013. [Article \(CrossRef Link\)](#)



**Ningning Hu** received the B.S. degree in Electronic Information Engineering in 2015 from the School of Information Science and Engineering, YanShan University. She is currently pursuing the M.S. degree in the Institute of Information Science, Beijing Jiaotong University. She works in computer vision and video compression. Her current research interest is view synthesis.



**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.



**Huihui Bai** received her B.S. degree from Beijing Jiaotong University, China, in 2001, and her Ph.D. degree from Beijing Jiaotong University, China, in 2008. She is currently a professor in Beijing Jiaotong University. 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).