COMBINED IMAGE PLUS DEPTH SEAM CARVING FOR MULTIVIEW 3D IMAGES

Vikas Ramachandra, Matthias Zwicker and Truong Q. Nguyen

CALIT2, University of California, San Diego, CA 92093

ABSTRACT

Multiview 3D displays have to multiplex a set of views on a single LCD panel. Due to this, each view has to be downsampled by a considerable amount leading to loss of details. In this paper, we extend the seam carving technique for adaptive resizing of images. It is proposed that the depth information be used along with the image pixel intensity values for resizing. This results in better resized multiview images. It is clear from the results presented that the object structure is maintained when the proposed method is used as compared to vanilla seam carving.

Index Terms— Image resizing, seam carving, depth map, freeview television, multiview, 3D.

1. INTRODUCTION

Multiview 3D (automultiscopic) displays [2] show 3D images and videos that can be viewed without special glasses. Further, by using more than 2 views, multiview 3D displays incorporate motion parallax, which gives the viewer more freedom, i.e. he does not need to keep his head position fixed. Also, the user can be present anywhere in the reasonably wide viewing zone. Typically, multiview displays multiplex 8, 16 or more number of views on the same LCD screen matrix. A large number of views provides a more realistic immersive viewing experience. However, due to the limited fixed resolution of the LCD display grid, each of the individual views must be resized before the spatial multiplexing process. At present, multiview displays only provide horizontal parallax, hence the views have to be downsampled along the columns only. For a fixed display grid resolution, the downsampling factor for each view is proportional to the number of views to be multiplexed [3]. Traditional sampling theory dictates that the views be suitably prefiltered before downsampling each of them. Recent advances in the field of content aware image resizing go beyond traditional sampling theory. The basic idea behind these techniques is that resizing of images should not only use geometric constraints, but consider the image content as well [[1], [5], [6], [7]]. One such image resizing technique which we use in this paper is called seam carving [1]. A seam is an optimal 8-connected path of pixels on a single image from top to bottom, or left to right, where optimality is defined by an image energy function. By repeatedly carving out seams in one direction we can change the aspect ratio or resize an image. This results in the most interesting regions of the image being retained, i.e. different regions of the image are affected differently. In this paper, an extension to seam carving suitable for multiview images is proposed. The multiview image format we concentrate on is the '2D plus depth' standard MPEG 3D format [4], which consists of one view and an estimated depth map using which other views can be generated. We show that using the depth map values along with the pixel intensity values gives superior resizing results than vanilla seam carving which uses only pixel values. This paper is organized as follows. In section 2, we briefly summarize seam carving. Section 3 contains a discussion of the multiview image format and the proposed resizing scheme. This section also contains a discussion about extending the idea to 3D video streams. This is followed by results in section 4 where we compare seam carving with proposed depth assisted seam carving. Section 5 concludes the paper.

2. SEAM CARVING

This section contains a discussion of seam carving condensed from [1]. A vertical seam is an 8-connected path of pixels in the image from top to bottom, containing one, and only one, pixel in each row of the image. Formally, let I be an nm image and define a vertical seam to be:

$$s^{x} = (s_{i}^{x})_{i=1}^{n} = (x(i), i)_{i=1}^{n}, s.t.\forall i, |x(i) - x(i-1)| \le 1$$
(1)

where x is a mapping $x : [1, \dots, n] \to [1, \dots, m]$. The pixels of the path of seam s (e.g. vertical seam) will therefore be: Removing the pixels of a seam from an image has only a local effect: all the pixels of the image are shifted left to compensate for the missing path. Given an energy function e, we can define the cost of a seam as: $E(s) = E(I_s) = \sum_{i=1}^{n} e(I(s_i))$. We look for the optimal seam s^* that minimizes this seam cost:

$$s^* = min_s E(s) = min_s \sum_{i=1}^n e(I(s_i))$$
 (2)

The optimal seam can be found using dynamic programming. The energy function used here was the L1 norm of the gradient, which can be defined as:

$$e_1(I) = |\partial I/\partial x| + |\partial I/\partial y| \tag{3}$$



Fig. 1. 2D image plus depth format examples.

3. 2D PLUS DEPTH IMAGE FORMAT AND PROPOSED RESIZING

Recently, MPEG has finalized the draft for coding videos for free viewpoint television as well as 3DTV. The selected compression technique deals with coding one of the views as well as an associated depth map. Examples of the image plus depth map set (for the Philips display) are shown in figure 1. For multiview displays, the depth map values are used (along with the existing reference view) to generate novel views. The modified seam carving algorithm is as follows. From, the depth map, a scaled inverse depth image is obtained. This image is appended to the existing image of the 2D view image. Then, seam carving is performed on the combined 2D image and depth signal. In this case, a vertical seam can be defined as:

$$s^{x} = (s_{i}^{x})_{i=1}^{n} = (x(i), i, d(i))_{i=1}^{n}, s.t. \forall i, |x(i) - x(i-1)| \le 1$$
(4)

where d(i) is the scaled inverse depth map value for the pixel at location x(i). The reason why seam carving is expected to perform better for 2D image plus depth input input rather than just with a 2D image input or just a depth map alone is as follows. A good depth map provides valid regions of gradient change *between* objects, i.e. inter object boundaries. However, the depth map may not identify some valid intra object gradients and boundaries. On the other hand, gradient detection for the 2D image identifies both valid and spurious boundaries at both intra and inter object regions. Combining the depth map with the 2D image helps provide a good competition between the intra object boundaries and valid inter object boundaries when the seams are being carved out. Since the L1 norm of the image gradient is used as the energy function for seam carving, combining 2D image and depth map in turn provides truer estimates of the energy function and thus improves the performance of the seam carving algorithm. Mathematically, this tradeoff between the inter object boundaries and intra object boundaries is incorporated into the energy function by the proposed method. Thus, the revised energy function from equation 3 can be written as:

where e'_1 is the modified energy function, α_{intra} and α_{inter} are the within object and between object importance weighting terms. D is the inverse depth map. It should be noted that the first term, in reality captures both inter and intra object gradients, therefore the α_s must be suitably adjusted to give comparable weighting to inter and intra object boundaries (Eg: $\alpha_{intra} = 0.75, \alpha_{inter} = 0.25$).

3.1. Extending the proposed method to 3D multiview videos

So far, our discussion has concentrated on multiview images. Here, we look at how the proposed depth assisted seam carving can be extended to multiview 3D videos. The basic idea is to introduce a temporal consistency measure for both the depth map stream as well as the video stream. For a given (multiview) video frame F_j , we only consider the immediate previous adjacent frame (F_{i-1}) , as was done in [5]. We motion compensate the adjacent frame F_{j-1} using the frame F_j as the reference, to obtain frame F'_{j-1} . One important modification is in the input to the motion estimation algorithm is that all the frames F_j, F_{j-1} actually contain the 2D image plus depth information, and are not just pixel intensity planes. Any true motion estimation scheme will suffice, we use a unidirectional block based motion estimator along with motion vector smoothing and refinement steps. Such true motion estimation ideas have been studied extensively in many papers (for example in [8], [9], [10]) and we will not reproduce those details here.

Since we are enforcing temporal smoothness constraints, the energy function for the k^{th} seam to be carved will now depend on all the k^{th} intermediate seam carved image (with k-1 seams removed) of the previous frame. Incorporating the adjacent frame constraints into the energy function, we can rewrite the energy function for the k^{th} seam being carved, as follows:

$$e_{1}''(I,k) = \alpha_{intra}(|\partial I/\partial x| + |\partial I/\partial y|) + \alpha_{inter}(|\partial D/\partial x| + |\partial D/\partial y|) + \alpha_{temp}|F_{j}^{k} - F_{j-1}^{k}|$$
(6)

where α_{temp} is the temporal smoothness weighting factor, and F_j^k and F_{j-1}^k are the intermediate images (2D plus depth combined) of the j^{th} and $(j-1)^{th}$ frames respectively.

4. RESULTS

In this section, we present 4 sets of images and their respective depth maps. For each, we apply seam carving on the intensity values. We also apply the proposed depth assisted seam carving method on the 2D image plus depth signal. For the figures with the input image signals (figures 2,4, 6 and 8), note that the left halves contain the intensity map and the right halves contain the depth maps. Also, for the fig-

$$e'_{1}(I) = \alpha_{intra}(|\partial I/\partial x| + |\partial I/\partial y|) + \alpha_{inter}(|\partial D/\partial x| + |\partial D/\partial y|)$$
 with the resized output images (figures 3,5, 7 and 9), the (5) left halves contain the vanilla seam carving results whereas

the right halves contain the results of the proposed method. Images 1 and 2 (figures 2 and 4) were originally 256x256 images, 100 columns were seam carved from each image to get a final output image size of 256x156 (figures 3 and 5). Images 3 and 4 (figures 6 and 8) were originally 800x640 images, 200 columns were seam carved from each image to get a final output image size of 800x440 (figures 7 and 9). From the results, it is clear that using the depth information helps retain the objects in a more coherent manner.

We also conducted a subjective visual test with these 4 sets of images. 4 image pairs (Each pair consisted of one image each for proposed method and seam carving) were presented to 10 viewers who was asked to judge if the left image was better than the right image and vice versa. The left image could either be the proposed method or vanilla seam carving and vice versa (i.e. the input order in the pair was randomized). The results of the test is presented in the histogram in figure 10. The negative valued bins show the viewer bias towards vanilla seam carving and positive valued bins indicate user preference to the proposed method. It is clear from the histogram that the viewers mostly prefer the proposed method's output (88 % of the time). This corroborates our expectation that depth information is beneficial for content aware resizing.



Fig. 2. Image 1: 2D and depth input.

5. CONCLUSION

In this paper, we have extended the content aware seam carving technique for adaptive resizing of images to multiview images. We have shown that the method performs well for the 2D image plus depth format. We also presented a discussion about extending this idea to multiview 3D video streams. Future work would include extending this idea to jointly resize a set of views, i.e. the extended H.264 MVC format. Also, more analysis is needed for 3D video streams to ensure temporal consistency of the seam carved multiview output.

6. REFERENCES

[1] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM SIGGRAPH 2007*, San Diego, 2007.



Fig. 3. Image 1, resized output: Seam carved (left), proposed (right).



Fig. 4. Image 2: 2D and depth input.



Fig. 5. Image 2, resized output: Seam carved (left), proposed (right).

- [2] C van Berkel, DW Parker and AR Franklin, "Multiview 3D LCD," *Proceedings of SPIE*, 1996.
- [3] J. Konrad and P. Agniel, "Subsampling models and anti-alias filters for 3-D automultiscopic displays," *IEEE Trans. Image Processing*, vol. 15, pp. 128-140, Jan. 2006.
- [4] A Bourge, J Gobert and F Bruls, "MPEG-C Part 3: Enabling



Fig. 6. Image 3: 2D and depth input.



Fig. 7. Image 3, resized output: Seam carved (left), proposed (right).





the introduction of video plus depth contents," *Content generation and coding for 3D-television workshop*, 2006.

[5] L Wolf, M. Guttmann and D. Cohen-Or, "Non-homogeneous Content-driven Video-retargeting," *IEEE 11th International Conference on Computer Vision*, Oct. 2007.



Fig. 9. Image 4, resized output: Seam carved (left), proposed (right).



Fig. 10. Histogram of the viewer preferences, positive bins indicate a leaning towards proposed method versus seam carving

- [6] D Simakov, Y Caspi, I Rehovot, E Shechtman and M. Irani, "Summarizing Visual Data Using Bidirectional Similarity," *IEEE Computer Vision and Pattern Recognition Conference*, 2008.
- [7] R Samadani, T Mauer, D Berfanger, J Clark and B Bausk, "Representative Image Thumbnails: Automatic and Manual" *Human Vision and Electronic Imaging Conference*, San Jose, Jan 2008.
- [8] C Stiller and J Konrad, "Estimating motion in image sequences" *IEEE Signal Processing Magazine*, 1999.
- [9] S Lee, O Kwon and R Park, "Weighted-adaptive motioncompensated frame rate up-conversion," *IEEE Trans. on Consumer Electronics*, vol.49, no.3, pp. 485-492, Aug. 2003.
- [10] G. Dane and TQ Nguyen, "Motion vector processing for frame rate up conversion," *IEEE International Conference on Acoustics, Speech, and Signal Processing,* May 2004.