Schwarz, S.; Sjöström, M.; Olsson, R., "Improved Edge Detection for EWOC Depth Upscaling," *Systems, Signals and Image Processing (IWSSIP), 19th International Conference on, 2012*, 11-13 April 2012

IMPROVED EDGE DETECTION FOR EWOC DEPTH UPSCALING

Sebastian Schwarz, Mårten Sjöström, and Roger Olsson

Department of Information Technology and Media, Mid Sweden University
Holmgatan 10, 85170, Sundsvall, Sweden
phone: + (46) 060 1488-36, fax: + (46) 060 1488-30,
email: [sebastian.schwarz, marten.sjostrom, roger.olsson]@miun.se

ABSTRACT

The need for accurate depth information in three-dimensional television (3DTV) encourages the use of range sensors, i.e. time-of-flight (ToF) cameras. Since these sensors provide only limited spatial resolution compared to modern high resolution image sensors, upscaling methods are much needed. Typical depth upscaling algorithms fuse low resolution depth information with appropriate high resolution texture frames, taking advantage of the additional texture information in the upscaling process. We recently introduced a promising upscaling method, utilizing edge information from the texture frame to upscale low resolution depth maps. This paper examines how a more thorough edge detection can be achieved by investigating different edge detection sources, such as intensity, color spaces and difference signals. Our findings show that a combination of sources based on the perceptual qualities of the human visual system (HVS) leads to slightly improved results. On the other hand these improvements imply a more complex edge detection.

***Index Terms*—** 3DTV, EWOC, depth map, ToF, upscaling, perceptual edge detection, HVS, CIE2000

1. INTRODUCTION

Sensor-fusion algorithms in video technology rely on feature information from one image source to adapt the resolution of a second source. These features, like edges and borders, can be extracted by edge-filtering methods, like the Canny edge detector [1]. The question is how to select the best combination of filters and color-spaces to achieve a most accurate scene description.

Autostereoscopic three-dimensional television (3DTV) is one of the major topics in consumer electronics in the last years. Depth-image-based rendering (DIBR) methods are used to reduce the transmitted data. This depth information can be obtained by correspondence matching between two or more reference views. Unfortunately these matching algorithms suffer from errors in occluded or texture-less regions. These errors can be reduced with dedicated range sensors, i.e. time-of-flight (ToF) cameras. The limited resolution of these sensors, compared to modern high-definition (HD) video, motivates the search for innovative upscaling algorithms, to fuse low resolution depth data with high resolution video frames.

There have been many proposals for sensor-fusion upscaling methods in recent years. Diebel et al. proposed a method based on Markov Random Fields (MRF [2]) and Kopf et al. introduced the well received joint-bilateral upscaling (JBU [3]). Most papers in this field rely on adaption of JBU, like the noise-aware filter for depth upsampling (NAFDU [4]) from Chan et al. or the pixel-weighted average approach (PWAS [5]), based on a ToF credibility map, from Garica et al.

Recently we introduced a novel depth-upscaling approach, solving an edge-weighted optimization problem [6]. Unlike above approaches we see the low resolution ToF data as a sparse representation of a full, i.e. high definition (HD) video resolution, depth map. We fill the missing values solving a linear least square problem weighted with edge information from the full resolution video frame as well as the low resolution ToF depth. The outcome of our Edge Weighted Optimization Concept (EWOC) is highly dependent on the results of the edge detection.

In the present paper we take a close look on which filters and color-spaces guarantee an accurate and thorough edge detection, leading to the best visual quality in view synthesis. We show that a color space combination based on the human visual system (HVS) leads to improved results. Having said that, these improvements are marginal and come at the cost of added complexity. We therefore present a simplified edge detection for EWOC with still competitive results.

The remainder of this paper is organized as follows: First we define the scope of this paper in Sec. 2. In Sec. 3 we give an overview of EWOC and color-edge detection in general, followed by our test arrangements in Sec. 4. Finally we present the results in Sec. 5 and conclude our work in Sec. 6.

2. PROBLEM STATEMENT

We see this work in the scope of sensor-fusion for 3DTV. In [6] we presented the capability of EWOC as depth upscaling method and its advantage to traditional (depth) upscaling.

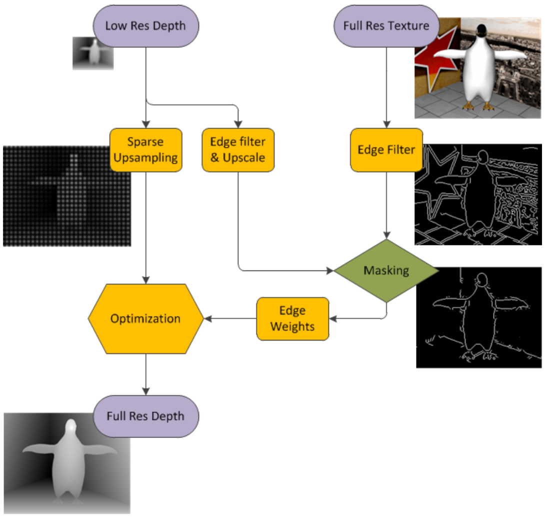


Fig. 1. EWOC: Low resolution depth upscaling with edge-weighted optimization.

We gave a detailed comparison to other upscaling methods such as JBU, PWAS and NAFDU and therefore exclude further comparisons from this paper. Instead, we concentrate on the influence of the edge map E_I, especially the effect of varied color edge detection, on the upscaling results.

3. UNDERLYING WORK

In order to understand the underlying work of this paper, we like to give an overview of the some fundamentals: EWOC, color edge detection and mean shift filter pre-processing.

3.1. Edge Weighted Optimization Concept (EWOC)

Fig. 1 shows the basic principle of the EWOC algorithm: Starting from a low resolution ToF depth map, we plot the known values on the corresponding positions of a full resolution depth map having the same resolution as the texture image, i.e. from a video source, resulting in a sparse depth map. The full resolution texture is edge filtered and the resulting edge map is masked with edge information from the low resolution ToF depth map. We fill the sparse depth map by solving a least square error problem using the masked edge map as weight: The spatial smoothness requirements in Eq. 1 and 2 encourage the depth of each pixel $d(x, y)$ to be similar to its spatial neighbors. To avoid depth blending at object borders, we introduce a weighting map Q_E that allows pixels on texture edges to be less similar.

$$Q_E(x, y) \cdot (d(x, y) - d(x + 1, y)) = 0 \qquad (1)$$

$$Q_E(x, y) \cdot (d(x, y) - d(x, y + 1)) = 0 \qquad (2)$$

The weighting map Q_E is generated from two parts defined in the following: One is the full resolution image I and the other is a mask gained from the edge information in the low resolution depth map D_{low}. We apply a combination of edge detectors on image I resulting in edge map E_I with a continuous value range of [0, 1]. E_I still contains many edges at no actual depth changes. Since this would lead to unwanted artifacts, we upscale D_{low} with bilinear interpolation and extract the depth edge map E_D to mask out unnecessary edges in E_I, thus leading to Q_E:

$$Q_E(x, y) = 1 - (E_I(x, y) \cdot E_D(x, y)) \qquad (3)$$

Eq. 1 and 2 define an over-determined system of linear equations, where certain depth values $d(x, y)$ are known from the low resolution depth map while others are unknown but defined by the linear equations. We solve these equations by finding the least square error solution using a block-active method [7]. A more detailed explanation of the whole process can be found in our paper introducing EWOC [6].

3.2. Color Edge Detection

Edge detection is one of the key applications in computer vision [1]. Typically edges were modeled as intensity discontinuities in grayscale images. Standard color edge detection expands this model on all the channels of a color space individually, simply combining the results. A good overview of edge detectors on multiple color spaces can be found in [8]. A more sophisticated approach is to calculate the color difference, the euclidean distance between two color vectors, and use its gradient for edge detection. Used on an uniform color space, such as CIELab, this results in an edge detection close to the human color difference perception. The CIE2000 color-difference formula was specially developed to represent the HVS [9] and is therefore predestinated for a perceptual color edge detection [10]. In this paper we apply a horizontal and vertical sobel filter on the CIE2000 color difference, combine and thin the resulting edges with non-maxima suppression and perform a hysteresis thresholding to connect 'broken' edges. We call the resulting edge map E_{CIE}.

3.3. Mean Shift Filtering

Another possibility to improve the edge detection is to remove unnecessary details, i.e. by smoothing the image with an edge-preserving mean shift filter. Thus making it easier to detect important edges. The mean shift filter was first introduced by Fukunaga and Hostetler [11] and is widely used in computer vision for image segmentation and clustering [12]. The filter clusters pixel within a defined spatial radius and color difference, calculates the color mean and assigns it to all pixel within the cluster. Several iterations are performed until the color mean remains constant. This results in a smoothened image with less overall color entities, accentuating (color) edges.

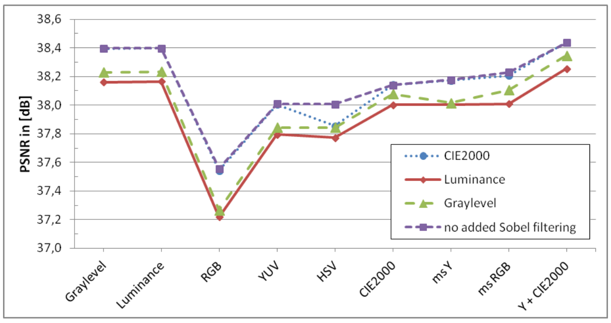
4. METHODOLOGY

The previous implementation of EWOC generated the edge map E_I from a combination of "Canny" edge detectors. The intensity edges were gained from the luminance channel Y. For the color edges we transfered image I to the HSV color space (Hue, Saturation and Value of brightness) and took the edges of each channel. We also added the Sobel filtered image I_G (grayscale version of I) to give E_I a continuous value range. The edge map E_I of image I can be described as follows:

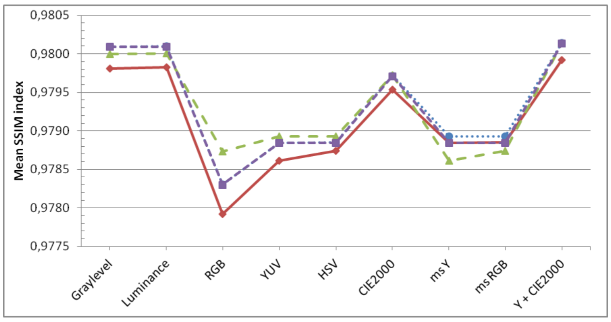
$$E_I = (C_Y \cup C_H \cup C_S \cup C_V) + (\frac{I_G * S_x}{255} + \frac{I_G * S_y}{255}) \quad (4)$$

Where C is the Canny result on the different color channels. S_x and S_y are the results of the horizontal and vertical Sobel operator respectively, which were only added at logical zeros to give a continuous value range of [0, 1]. For this paper we take a look at color edge detection in different color spaces and also the possible enhancement by perceptual color edge detection or pre-filtering.
The following sources for edge detection have been evaluated: For "intensity only" edges we took the luminance channel of YUV and a grayscale version of RGB. For color edge detection we combined the single channels of the color spaces RGB, YUV and HSV. For a perceptual color edge detection we used the CIE2000 color difference as described in 3.2. To evaluate the possible effects of pre-processing with a mean shift filter (spatial radius: 4, color difference: 4.0) we used the luminance channel and the combined RGB color space, for both intensity and color evaluation. Another possible scenario for an improved edge detection is the combination of the edge maps for intensity and color difference. Therefore we combined E_Y and E_{CIE} into a single depth map. E_{CIE} was gained as described in Sec. 3.2, but before the hysteresis thresholding we add E_Y and apply the thresholding on the combined results. Finally we also varied the continuous values of E_I by taking different sources for the Sobel filtering: Grayscale, luminance (Y channel), and CIE2000 color difference as well as no added Sobel filtering.

Our ToF capture system is based on a Fotonic C-70[13] ToF camera, providing a 160x120 pixel resolution, combined with a 1280x960 pixel machine vision camera. This leads to an image-to-depth ratio of 64:1, equivalent to a subsampling factor of 8 in the x- and y-directions, respectively. To objectively evaluate the proposed solution we considered pseudo ToF data, i.e. subsampled ground truth depth. We used a set of different test sequences, consisting of both computer generated scenes with synthetic ground truth depth maps as well as realistic footage with estimated depth maps. Our results are similar for all sequences, but to limit the extent of this paper, we will only present the test sequence "Poznan Street", a realistic 1920x1088 pixel resolution sequence with estimated depth maps [14]. We choose this sequence because of its high resolution and its open availability for the research community.
The overall aim of depth map upscaling is a similar view synthesis quality compared to syntheses using full resolution depth maps. Thus we compared syntheses using upscaled depth maps to syntheses using the given full resolution depth maps, eliminating errors introduced from the view synthesis algorithm from our evaluation. The evaluation criteria used are peak signal-to-noise ratio (PSNR), as the standard similarity measure in image quality assessment, and the structural similarity index (SSIM), to address the special characteristics of the HVS.



(a) PSNR for synthesis with upscaled depth compared to true view



(b) SSIM index for synthesis with upscaled depth compared to true view

Fig. 2. PSNR & SSIM index comparison for different edge detection sources. Average of 20 frames for view 4 of test sequence "Poznan Street". Upscaling factor 8. "ms" stands for mean-shift filtered values, "Y" for luminance values.

5. RESULTS

Our results in Fig. 2 show several things: Firstly the best results are achieved with an "intensity only" edge detection, i.e. graylevel & luminance. This was to be expected since the human eye is most sensitive to intensity changes. Secondly the added Sobel filter for continuous edge values has little effect on the resulting view synthesis. If it comes to color edge detection the CIE2000 color difference seems the only source able to compete with intensity signals, due to its close resemblance of the HVS. We also show that pre-processing with a mean shift filter might be beneficial for image segmentation or object recognition, but does not lead to an improvement in sensor-fusion. Finally, the combination of intensity and color difference gives the best quality in view syntheses.

This quality is gained at the cost of higher complexity. The necessary steps to gain E_{CIE} increases the processing time on edge detection by around 150 times compared to a simple Canny edge detector on the Luminance channel. Usually edge detection takes only a small portion of the overall processing time budget (around 2%). With the CIE2000 color difference this portion is increased to almost 75%.

6. CONCLUSION

We looked at various sources for a thorough (color) edge detection for EWOC depth map upscaling, aiming at a higher quality in view syntheses. Popular approaches in image segmentation, i.e. mean shift filtering, were found less adequate in this sensor-fusion application. Perceptual color representation, i.e. the CIE2000 color difference, can provide improvements if combined with intensity information. On the other hand these improvements are small compared to the increase in complexity. Generally, an intensity signal is sufficient to achieve respectable upscaling results. We therefore reduce the edge detection algorithm to the luminance channel. Instead of the intended increase in quality we found a decrease in complexity. In future research we will further reduce complexity, aiming at a real-time EWOC depth map upscaling for ToF sensors. This may include more upcoming edge detectors.

Acknowledgments

This work has been supported by grant 2009/0264 of the KK Foundation, Sweden, by grant 00156702 of the EU Objective 2 programme, Mellersta Norrland, Sweden, and by grant 00155148 of local government Västernorrland, Sweden.

7. REFERENCES

[1] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679–698, November 1986.

[2] J. Diebel and S. Thrun, "An application of markov random fields to range sensing," in *Proceedings of Conference on Neural Information Processing Systems*, Cambridge, MA, 2005, MIT Press.

[3] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Transactions on Graphics*, vol. 26, no. 3, 2007.

[4] D. Chan, H. Buisman, C. Theobalt, and S. Thrun, "A noiseaware filter for real-time depth upsampling," in *Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications*, 2008.

[5] F. Garcia, B. Mirbach, B. Ottersten, F. Grandidier, and A. Cuesta, "Pixel weighted average strategy for depth sensor data fusion," in *IEEE 17th International Conference on Image Processing*, 2010.

[6] S. Schwarz, M. Sjöström, and R. Olsson, "Depth map upscaling through edge weighted optimization," in *3D Image Processing (3DIP) and Applications*, 2012.

[7] L. F. Portugal, J. J. Júdice, and L. N. Vicente, "A comparison of block pivoting and interior-point algorithms for linear least squares problems with nonnegative variables," *Mathematics of Computation*, vol. 63, pp. 625–643, October 1994.

[8] S. Wesolkowski, M.E. Jernigan, and R.D. Dony, "Comparison of color image edge detectors in multiple color spaces," in *Image Processing, 2000. Proceedings. 2000 International Conference on*, sept. 2000, vol. 2, pp. 796 –799 vol.2.

[9] M. R. Luo, G. Cui, and B. Rigg, "The development of the cie 2000 colour-difference formula: Ciede2000," *Color Research & Application*, vol. 26, no. 5, pp. 340–350, 2001.

[10] Li Xue-Wei and Zhang Xin-Rong, "A perceptual color edge detection algorithm," in *Computer Science and Software Engineering, 2008 International Conference on*, dec. 2008, vol. 1, pp. 297 –300.

[11] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *Information Theory, IEEE Transactions on*, vol. 21, no. 1, pp. 32 – 40, jan 1975.

[12] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 5, pp. 603 –619, may 2002.

[13] Fotonic, "C70 time-of-flight camera," http://www.fotonic.com/.

[14] M. Domañski, T. Grajek, K. Klimaszewski, M. Kurc, O. Stankiewicz, J. Stankowski, and K. Wegner, "Poznañ multiview video test sequences and camera parameters," ISO/IEC JTC1/SC29/WG11 MPEG 2009/M17050, October 2009, Xian, China.