MODIFIED WEAK FUSION MODEL FOR DEPTHLESS STREAMING OF 3D VIDEOS

Dogancan Temel, Qiongjie Lin, Guangcong Zhang and Ghassan AlRegib

School of Electrical and Computer Engineering, Georgia Institute of Technology
Atlanta, GA, 30332-0250 USA
{cantemel, lin3, zhanggc, alregib} @ gatech.edu

ABSTRACT

We propose a nonlinear fusion model to optimally reconstruct depth maps from monocular cues. The authors in [1] showed the effectiveness of estimating depth from individual cues, which saves bandwidth without compromising the quality. In this paper, we combine such depth map estimates as a Sensor Fusion problem. These depth map estimates are extracted from color, motion and texture structure of the scene. We address the combination of monocular cues as a nonlinear optimization problem with linear constraints. 2D and 3D objective quality metrics are used as reliability metrics. At first, we perform Particle Swarm Optimization based on PSNR to obtain an initial estimate for the linear weights. Then, we use these weights to perform Active-Sets based on 3VQM. We tested our approach on various video sequences with different monocular characteristics. Experimental results show that the quality of the rendered views based on combined depth maps is comparable to the quality of rendered views based on ground truth depth maps while the savings in bandwidth is up to 38.8%.

Index Terms— Monocular cues, 3DTV, sensor fusion, non-linear optimization

1. INTRODUCTION

Recently, three-dimensional television (*3DTV*) and free viewpoint television (FTV) have attracted great attentions in both academia and industry. This technology enables the users to navigate through the scene with a desired viewpoint. A typical system structure of *3DTV* processing chain is composed of six main components: (*i*) 3D video capturing and content generation; (*ii*) coding 3D video; (*iii*) transmitting; (*iv*) decoding the received sequences; (*v*) generating virtual views; (*vi*) displaying the stereoscopic images on the screen.

The simplest way to broadcast 3DTV is to send right and left views separately so that we can synthesize the virtual views via stereo matching methods. However, such method requires require twice the bandwidth compared to *2D* video streaming. Instead of stereo, depth is utilized where a single colored image and a depth map for each frame are transmitted. On average, depth map requires one fifth the bandwidth required by a colored image [2]. Therefore, depth-based methods lead to significant bandwidth savings besides the elimination of discomfort sources while generating virtual views. Among all view synthesis techniques in the literature, Depth Image Based Rendering (DIBR) is one of the most commonly used ones [2, 3]. *Occlusion-disocclusion* is the main problem when DIBR is used, which leads to *holes (unassigned black pixels)* in the synthesized views. State-of-the-art methods that are used for hole filling are investigated in [4]. The quality of the synthesized views is highly correlated with the depth maps used for rendering. A reliable depth map is essential to obtain a high quality of experience.

In video-plus-depth systems, video and depth are streamed over two different channels. Both are compressed using H.264, for example. As shown in the literature, depth map processing degrades the quality of the rendered views due to quantization, compression and transmission errors [5]. An alternative approach is to eliminate the depth channel and instead send depth cues with the video reference channel [1]. In this case, source side is assumed to have a reference video and the corresponding depth map video. Based on the given depth map, a number of depth cues per frame are generated and sent with the corresponding video reference frame without requiring the transmission of the complete depth map. At the receiver side, an algorithm is run to use monocular cues in the video, e.g., luminance, chrominance, motion, and texture, and the received depth cues to reconstruct the complete depth map for a particular frame. Although this approach showed some promising initial results, the combination of all depth estimates from all cues remains as a challenge.

In the presence of multiple depth estimates, depth maps are used according to the reliability of each estimate. Instead of using the estimates with highest reliability, integrating all of the estimates turned out to be a promising approach. In [1], a brute force approach was presented to solve the combination problem. However, authors in [1] did not focus on the optimal combination. Combining depth estimates can be considered as a nonlinear optimization problem because of the non-linear objective functions used for optimization. In general, optimization problems can be classified into two groups as global and local optimization. *Genetic Algorithm*, *Neural Network Model* and *Firefly Algorithm* are commonly used for global optimization whereas *Trust Region* and *Active-Set* methods are preferred for local optimization in the literature. Detailed information about numerical optimization methods is provided in [6]. In order to perform an optimization, we need to define an objective function that is correlated with the quality of synthesized views. The majority of the objective *3D* video quality metrics in the literature are basically extensions of existing *2D* metrics. The correlation between objective quality metrics and subjective results for video-plus-depth-based stereoscopic videos is investigated in [5].

In this paper, we propose a sensor fusion model based on non-linear optimization to combine estimated depth maps, and reconstruct the rendered views at the receiver side. To achieve this goal, we propose a feedback-based *Modified Weak Fusion (MWF)* model. The rest of the paper is organized as follows: In section 2, *MWF* model is established for combining estimated depth maps and the combination process is modeled as a non-linear optimization problem with linear constraints. Experimental results are discussed in Section 3. Conclusion and future directions are given in Section 4.

2. MODIFIED WEAK FUSION MODEL FOR COMBINING ESTIMATED DEPTH MAPS

The main contribution of the proposed work is finding an optimal combination of cues for the reconstruction of depth map. The authors of [7] proposed the reconstruction of depth maps from monocular cues based on color, motion and texture and they followed a brute-force approach to combine all depth cues in [1].

Instead of transmitting depth with the reference video, we only send depth cues as a side information with the reference video. At the transmitter, we generate the depth cues based on the ground truth depth. However, we do not actually need to use the ground truth depth. For our purpose, we only need an initial depth estimate, which can be obtained by adding a depth estimation step at the transmitter side. Our system instead focuses on reconstructing a high quality depth map at the receiver using monocular and depth cues. We use the transmitted depth cues and the *2D* video to reconstruct the depth maps before the DIBR stage. We will briefly explain the monocular cues extraction and depth map reconstruction in the following paragraphs. However, readers are encouraged to read [7] that has the original idea of using monocular cues.

At the transmitter side, we calculate the histogram of the depth map and then we extract the depth planes that correspond to the distance where most of the content is concentrated in the scene. We send these extracted depth planes along with the color video to perform histogram matching and convert relative depth maps to absolute ones. After we run a number of histogram matching operations, in all the videos we experimented with, we end up with 4 depth levels that can describe the depth structure while mapping between depth estimates and depth cues. Nearest and furthest depth planes are transmitted as side information and used as monocular depth cues.

At the receiver side, we extract depth estimates from luminance, chrominance, motion and texture. In the luminance case, we calculate the histogram of the luminance component in the corresponding reference video frame. Then, we start a number of histogram matching operations that lead us to assign luminance regions from the luminance histogram to depth levels. The same process is repeated for chrominance. In the motion and texture case, we developed and implemented an algorithm that jointly maps motion and texture to depth levels.

After extraction, we need to combine monocular cues to obtain a single depth map for rendering the virtual view. Combining monocular cues can be considered as a sensor fusion problem. In general, fusion methods are classified into two groups as *Weak Fusion (WF)* and *Strong Fusion (SF)* [8]. *WF* assumes that depth cues are isolated and the final estimate is obtained by linearly combining the individual estimates. However, *SF* combines depth estimates as an interactive and holistic process. In the proposed work, we modeled integration of monocular cues as a *Feedback-based Modified Weak Fusion Model* as it is illustrated in Figure 1. Z_1 is the luminance based depth cue, Z_2 and Z_3 are the depth cues based on the chrominance, and Z_4 is the motion/texture based depth cue. The final depth map $\tilde{Z}$ is a linear combination of $\{Z_i\}_{i=1}^{4}$,

$$\tilde{Z} = \sum_{i=1}^{4} w_i \otimes Z_i \tag{1}$$

where $Z_i \in \mathbb{R}^{N \times M \times 1}$, $\{w_i\}_{i=1}^{4} \in \mathbb{R}$ is the linear combination weights and $\otimes$ means the pixel-wise product.

2.1. Indirect objective quality assessment for depth maps

We are interested in the quality of the rendered videos. However, objective quality metrics that are directly based on depth maps are not highly correlated with the subjective results. Thus, weight assignment for the combination of depth maps will be based on evaluating the

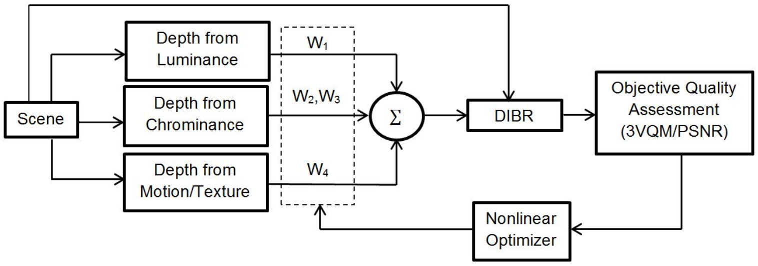


Fig. 1. Feedback-based Modified Weak Fusion Model

quality of the rendered views. In the proposed work, *PSNR* and *3VQM* are used as the objective quality metrics [9]. *PSNR-based Full-Reference Objective Quality Measure* is based on the 2D quality measure of the synthesized views. In our case, the evaluation is performed in a number of consecutive frames. Thus the *PSNR* based metric is defined as the mean value of the *PSNR* values over N frames:

$$E_P = -\frac{1}{N}\sum_{n=1}^{N} PSNR(I_n;\ I_n^{ideal}) \tag{2}$$

where I_n and I_n^{ideal} are DIBR-synthesized images based on the combined depth map $\tilde{Z}$ and the ground truth depth map. *PSNR* does not have a high correlation with subjective results. However, it is mainly used for initialization purposes.

3VQM-based Objective Quality Measure is based on a 3D Video quality metric that is used for synthesized views [9]. *3VQM* utilizes three distortion measures that evaluate the temporal and spatial variation of depth errors that would lead to inconsistencies between the left and right view, fast changing disparities, and geometric distortions. These measures are the spatial outliers (**SO**), temporal outliers (**TO**), and temporal inconsistencies (**TI**), defined as:

$$\mathbf{SO} = \mathbf{std}(\triangle Z) \tag{3}$$

$$\mathbf{TO} = \mathbf{std}(\triangle Z_{t+1} - \triangle Z_t) \tag{4}$$

$$\mathbf{TI} = \mathbf{std}(Z_{t+1} - Z_t) \tag{5}$$

where $\triangle Z = |Z_{ideal} - Z|$ is the difference between the ideal depth and the estimated depth; *std* is the standard deviation. These measures are pooled into one 3D vision-based quality metric for stereoscopic DIBR-based videos, which is *3VQM*:

$$\mathbf{3VQM} = K\,(1-\mathbf{SO}(\mathbf{SO}\cap\mathbf{TO}))^{a}\,(1-\mathbf{TI})^{b}(1-\mathbf{TO})^{c} \tag{6}$$

where K is a constant for scaling. If the evaluation is performed in N consecutive frames, the final 3VQM value is defined as the average of N 3VQM values.

2.2. Combination as a Nonlinear Optimization Problem with Linear Constraints

Our goal is to determine the optimal assignment of weights for monocular cues that results in highest quality w.r.t. the metrics mentioned in Section 2.1. The problem can be described as a nonlinear optimization model with linear constraints as follows:

$$\begin{aligned} &\underset{w_i}{\textbf{maximize}} && F_{3VQM}(\tilde{Z}(w_i)) \text{ OR } F_{PSNR}(\tilde{Z}(w_i)) \\ &\textbf{subject to} && w_i \in [0,1];\ \sum_{i=1}^{4} w_i = 1 \end{aligned} \tag{7}$$

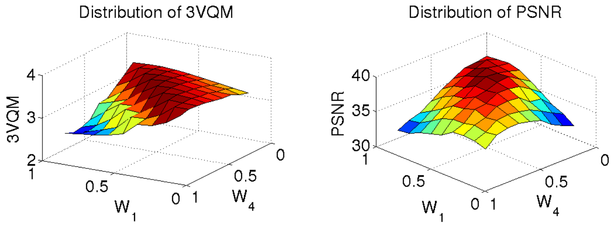
Nonlinear optimization model can be respectively adopted to *PSNR* and *3VQM* as follows:

$$\begin{aligned} &\underset{w_i}{\textbf{minimize}} && R = -\frac{1}{N}\sum_{n=1}^{N} PSNR(I_n;\ I_n^{ideal}) \\ &\textbf{subject to} && w_i \in [0,1];\ \sum_{i=1}^{4} w_i = 1 \end{aligned} \tag{8}$$

where $I = DIBR\left(\sum_{i=1}^{4} w_i \otimes Z_i\right)$.

$$\begin{aligned} &\underset{w_i}{\textbf{minimize}} && R = K - F_{3VQM}\left(\sum_{i=1}^{4} w_i \otimes Z_i\right) \\ &\textbf{subject to} && w_i \in [0,1];\ \sum_{i=1}^{4} w_i = 1 \end{aligned} \tag{9}$$

where $F_{3VQM}(Z)$ corresponds to the *3VQM* value of the synthesized view based on the depth map Z; the residual function $R = K - F_{3VQM} \in [0,K]$, because $\forall\ Z \in \mathbb{R}^{n\times m}$, $F_{3VQM}(Z) \in [0,K]$. w_1 is the weight for luminance (Y), w_2 and w_3 are for chrominance (Cr and Cb) and w_4 is for Motion-Texture.



(a) 3VQM

(b) PSNR

Fig. 2. 3VQM/PSNR distribution w.r.t. w_1 and w_4 where w_2=0; w_3=1-w_1-w_4 for the *Balloons* sequence.

2.3. Brute-Force Search

Brute-force search is performed for both *3VQM* and *PSNR* to determine the relationship between the weight assignment and the quality of the rendered view. Contours of *3VQM* and *PSNR* are depicted in Figures 2. In order to show the distribution patterns in a 3D plot, 0.0 is assigned to the weight of Cr channel and the weight of Cb is basically 1 minus the sum of other weights. In the following sections, we'll show that optimal weight of the Cr channel is negligible compared to others.

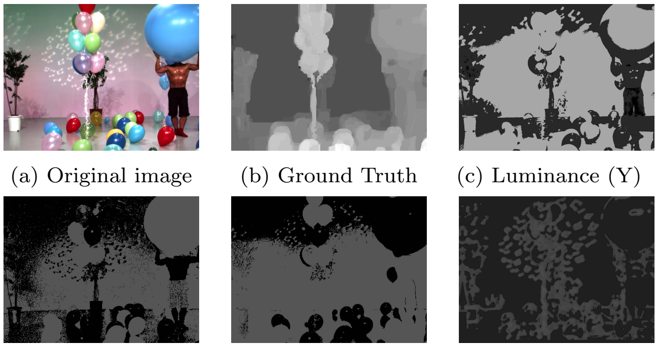
Our goal is to detect the similarities in the distribution patterns of *3VQM* and *PSNR* so that we can use *PSNR* instead of *3VQM* to minimize computational complexity. We will utilize the regions that correspond to high objective quality values such as the red peaks where both w_1 and w_4 are close to 0.5 as in Figure 2. These results indicate that luminance and motion-texture should significantly contribute to the optimal depth estimate of the *Balloons* sequence whereas weights of chrominance cues are significantly lower.

We can validate the assignment of weights by looking at the relative depth estimates in Figure 3. Relative depth based on luminance channel (3.c) distinguishes the difference between foreground and background for most of the regions since luminance represents the intensity channel of the scene which varies based on the distance w.r.t. the camera. However, chrominance channels are related to the color distribution of the scene. When we have a special scene configuration with black background and colorful foreground objects as in the *Balloons* sequence, relative depths based on chrominance channels may not be highly correlated with ground truth depth map. *Cr* channel (3.d) still contains reliable information for some local regions such as the upper part of the person and the plant whereas *Cb* (3.e) is mostly unreliable. Motion/texture combination captures the edges of the foreground objects that are in motion. Shades at the background lead to unreliable regions in the depth estimate however other regions provide reliable relative depth information. Therefore, weights of luminance (w_1) and motion-texture (w_4) are expected to be close to 0.5 as it was shown in Figure 2.

2.4. Active-Sets Based Nonlinear Optimization

Residual functions in the optimization problem are nonlinear as they are expressed in Eq. 8 and Eq. 9. Therefore, we can treat the combination problem as a nonlinear optimization with linear constraints. To simplify the problem and make the approach more feasible, we can linearize the nonlinear residual functions and approximate the solution in discrete cases.

(A) Linearizing Nonlinear Residual Functions



(a) Original image (b) Ground Truth (c) Luminance (Y)

(d) Chrominance (Cr) (e) Chrominance (Cb) (f) Motion-Texture

Fig. 3. Color image, ground truth depth-map and relative depth-maps for the `Balloons` sequence

As $F : \mathbb{R}^n \mapsto \mathbb{R}$, given a vector function $f : \mathbb{R}^n \mapsto \mathbb{R}^m$, formulation of nonlinear optimization problem is :

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}}\{F(\mathbf{x})\} \tag{10}$$

where

$$F(\mathbf{x}) = \frac{1}{2}\|\mathbf{f}(\mathbf{x})\|^2 = \frac{1}{2}\mathbf{f}(\mathbf{x})^T\mathbf{f}(\mathbf{x}) \tag{11}$$

Provided that f has continuous second partial derivatives, the residual function can be written in the linear form with Taylor expansion as:

$$\mathbf{f}(\mathbf{x}+\mathbf{h}) = \mathbf{f}(\mathbf{x}) + \mathbf{J}(\mathbf{x})\mathbf{h} + O(\|\mathbf{h}\|^2) \tag{12}$$

where J is the Jacobian. It yields:

$$\mathbf{F}'(\mathbf{x}) = \mathbf{J}(\mathbf{x})^T\mathbf{f}(\mathbf{x}) \tag{13}$$

$$\mathbf{F}''(\mathbf{x}) = \mathbf{J}(\mathbf{x})^T\mathbf{J}(\mathbf{x}) + \sum f_i(x)\mathbf{f}_\mathbf{i}''(\mathbf{x}) \tag{14}$$

and thus $F(x)$ can be linearized as:

$$F(\mathbf{x}+\mathbf{h}) \backsimeq F(\mathbf{x}) + \mathbf{h^T J^T f} + \frac{1}{2}\mathbf{h^T J^T J h} \tag{15}$$

Up to this point, we have transformed the nonlinear residual function into linear residual function, and the nonlinear programming turned into a quadratic programming (QP).

(B) Approximation in Discrete Cases

Discussions provided in *part A* are based on analytical results. However, we are not able to directly apply these expressions to our case since analytical expressions of the PSNR-based and 3VQM-based residual functions are not available. Therefore, we need to come up with an approximation which can be solved by replacing the *Jacobian* with central-difference:

$$\mathbf{J}(\mathbf{x}) = f(\mathbf{x} + \frac{1}{2}\delta) - f(\mathbf{x} - \frac{1}{2}\delta) \tag{16}$$

and replacing the *Hessian* matrix $\mathbf{H}(\mathbf{x}) = \mathbf{J}(\mathbf{x})^T\mathbf{J}(\mathbf{x})$ with each element of $\mathbf{H}$ can be expressed as:

$$H_{i,j}(x) = f(x+\delta_i+\delta_j) - f(x+\delta_j) - f(x+\delta_i) + f(x) \tag{17}$$

(C) **Active-set Methods for Constrained Quadratic Programming**

The residual functions of Eq. 8 and Eq. 9 can be written as a Linear Constrained QP as follows:

$$\begin{aligned} \underset{\mathbf{w}\in\mathbb{R}^4}{\text{minimize}} \quad & \tilde{R}(\mathbf{w},Z) \\ \text{subject to} \quad & A\mathbf{w} \geqslant \mathbf{0}_{4\times 1} \\ & B\mathbf{w} \geqslant [-1,-1,-1,-1]^T \\ & C^T\mathbf{w} = 1 \end{aligned} \tag{18}$$

where $\tilde{R}$ is the linearized quadratic residual function for *PSNR* and *3VQM* respectively; $A = \mathbf{I}_{4\times 4}$; $B = -\mathbf{I}_{4\times 4}$; $C = [1,1,1,1]^T$. This linear constrained *QP* problem can be solved by *Active-Sets (AS)* [6]. The advantages of this method come from fast convergence and high precision. However, *Active-Sets* can only find local optimal solutions since output depends on the initialization of weights.

2.5. Modified Weak Fusion Workflow

Due to the high computational complexity of *3VQM*, it is not feasible to directly search for the global optimal solution. However, as it is shown in Figure 2, distribution patterns of *PSNR* and *3VQM* are similar in specific regions. So, we can take advantage of the lower computational complexity of *PSNR*. We perform *Particle Swarm Optimization (PSO)* with *PSNR* to obtain initial weight assignments. Then, *Active Sets* algorithm can be initialized with these weights to find global maxima w.r.t. *3VQM*. Workflow for the AS-based optimization with PSO-based initialization is described in Figure 4, where $\{w_i\}_{i=1:4}$ are weights for depth map estimates.

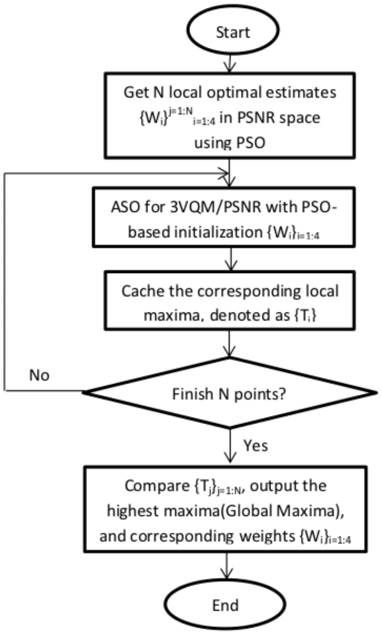

Start
Get N local optimal estimates $\{W_i\}^{j=1:N}_{i=1:4}$ in PSNR space using PSO
ASO for 3VQM/PSNR with PSO-based initialization $\{W_i\}_{i=1:4}$
Cache the corresponding local maxima, denoted as $\{T_i\}$
Finish N points?
No
Yes
Compare $\{T_j\}_{j=1:N}$, output the highest maxima(Global Maxima), and corresponding weights $\{W_i\}_{i=1:4}$
End

Fig. 4. Modified Weak Fusion Workflow

3. RESULTS AND ANALYSIS

3.1. Active Sets with Random Initialization

We randomly chose four sets of initial weights to perform *AS-based* optimization and the results are provided in Table 1. Optimization efficiency is highly sensitive to the initialization since we are using local optimizers. For the first two initialization sets, we observe a significant increase in the quality metrics, especially in *3VQM*. According to the simulation results of *Balloons* sequence, we can obtain increases up to *1.21 dB* in *PSNR* and *1.09* in *3VQM*. AS-based optimization assigns majority of the weights to Luminance (w_1) and Motion-Texture (w_4) whereas *Chrominance* channels *Cr* (w_2) and *Cb* (w_3) get lower weights. This tendency is more obvious when the objective function is *3VQM* as it is shown in Table 1. Assignment of higher weights to luminance and motion-texture is expected due to the higher reliability of these cues as it is explained in Section 2.3.

3.2. Active-Sets with PSO-based Initialization

We performed Particle Swarm Optimization (PSO) under varying number of iterations and *population of candidate solutions*. Active-Sets based nonlinear optimization is simply achieved by repeating steps in section 2.4 with initial weights assigned from the output of *PSNR-based PSO*. Experimental results in terms of *PSNR* and *3VQM* are shown in Table 2. Random initialization of weights can lead up to *26.95 dB* in *PSNR*. When we perform *Active-sets, PSNR* can be increased by *1.21 dB*. Initializing *Active-sets* with *PSO* leads to further increment in *PSNR* of about *1.8 dB*. In the case of *3VQM*, random initialization leads to *3.41* out of *5.00*, on average. When we perform *Active-sets*, *3VQM* increases up to *4.50*. *3VQM* can be further increased up to *4.76* by initializing *Active-sets* with *PSO*. Thus, we can obtain the highest objective quality results when we perform *Active Sets* with *PSO*-based initialization.

3.3. Quality-Bandwidth tradeoff for *MWF*

The proposed method eliminates the transmission of depth map and sends depth cues as a side information. Therefore, we need to compare between the proposed method and the *Two-Channels* approach where we compress the depth with *H.264*. We use *ver. 18.4* of *H.264/AVC* software to encode color and depth videos [10] . Entropy coding method is *CABAC* and *QP* is set to *40*. *3VQM* values of the synthesized videos based on compressed depth are similar to the rendered videos based on monocular cues. However, the *PSNR* values for the rendered views based on compressed depth maps are *6.4 dB* higher than the rendered videos based on reconstructed depth, on average. For the *Two-Channels* case, we compress both depth and color videos. However, in

Table 1. Active-Sets with Random Initialization

Initial Weights and Quality		Optimized Weights and Quality	
[w1,w2,w3,w4]	**PSNR**	**[w1,w2,w3,w4]**	**PSNR**
[0.25, 0.25, 0.25, 0.25]	**26.96**	[0.3088, 0.1149, 0.0753, 0.5010]	**28.16**
[0.30, 0.30, 0.30, 0.10]	**26.24**	[0.1862, 0.1963, 0.0741, 0.5434]	**28.09**
[0.40, 0.10, 0.10, 0.40]	**27.78**	[0.3104, 0.1129, 0.0711, 0.5056]	**28.15**
[0.45, 0.05, 0.05, 0.45]	**26.80**	[0.1516, 0.1108, 0.0459, 0.6917]	**27.81**
[w1,w2,w3,w4]	**3VQM**	**[w1,w2,w3,w4]**	**3VQM**
[0.25, 0.25, 0.25, 0.25]	**2.76**	[0.4345, 0.0100, 0.0479, 0.5076]	**4.67**
[0.30, 0.30, 0.30, 0.10]	**2.41**	[0.8989, 0.0100, 0.0811, 0.0100]	**4.60**
[0.40, 0.10, 0.10, 0.40]	**4.04**	[0.5316, 0.0100, 0.0100, 0.4484]	**4.64**
[0.45, 0.05, 0.05, 0.45]	**4.43**	[0.4254, 0.0100, 0.0677, 0.4969]	**4.67**

the proposed method, we only compress color video and send depth cues without compression. We only send the depth planes that are closest and furthest which result in *2 Bytes* side information per frame. Package size is around *156 kB* for the proposed method whereas *Two-Channels* approach leads to a package size of about *255 kB*. Proposed method leads to saving ratios up to *38.8 %* in bandwidth compared to the *Two-Channels* case for the *Balloons* sequence.

Table 2. *3VQM* and *PSNR* results. Random: Random weight assignment, ASO:Active Sets Optimization, PSO:Particle Swarm Optimization and Two-Channels: Color and depth videos are compressed with H.264.

	PSNR	**3VQM**
Random	26.95	3.41
ASO	28.16	4.50
ASO with PSO	29.96	4.76
Two-Channels	36.38	4.64

4. CONCLUSION

In this paper, we proposed a novel approach to integrate monocular depth cues as a sensor fusion problem. We approached the fusion problem as a nonlinear optimization with linear constraints. Since our objective quality metric was highly non-linear, we preferred the *Active-Sets (AS)* method to get fast convergence rate and high accuracy. To overcome the local maxima problem, we initialized *AS* with the weights obtained from *PSO*. Results show that *AS-based* optimization with *PSO-based* initialization increased *PSNR* by *3.01dB* and *3VQM* by *1.35* with respect to the random weight assignment. When the proposed method and *Two-Channels* approach is compared, *3VQM* values remain similar, *PSNR* decreases mildly and bandwidth savings increase significantly. In our future work, we plan to extend the types of monocular cues used in the integration process to guarantee a high quality of experience.

5. REFERENCES

[1] M. Aabed, D. Temel, M. Solh, and G. AlRegib, "Depth map estimation in dibr stereoscopic 3d videos using a combination of monocular cues," in *Asilomar'12*, Pacific Groove, CA, Nov. 2012.

[2] C. Fehn, "Depth-image-based Rendering (DIBR), Compression, And Transmission For A New Approach On 3DTV," *Proc. of SPIE*, vol. 5291, pp. 93–104, 2004.

[3] S. C. Chan, H. Shum, and K. Ng, "Image-based rendering and synthesis," *Signal Processing Magazine, IEEE*, vol. 24, no. 6, pp. 22 –33, Nov. 2007.

[4] C. Vázquez, W. J. Tam, and F. Speranza, "Stereoscopic imaging: filling disoccluded areas in depth image-based rendering," in *Three-Dimensional TV, Video, and Display V*, 2006.

[5] C. T. E. R. Hewage, S. T. Worrall, S. Dogan, S. Villette, and A. M. Kondoz, "Quality evaluation of color plus depth map-based stereoscopic video," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 3, no. 2, pp. 304 –318, Apr. 2009.

[6] J. Nocedal and S. J. Wright, "Numerical optimization," Berlin, New York: Springer-Verlag.

[7] D. Temel, M. Aabed, M. Solh, and G. AlRegib, "Efficient streaming of stereoscopic depth-based 3d videos," in *SPIE'13*, San Francisco, CA, Feb. 2013.

[8] M. S. Landy, L. T. Maloney, E. B. Johnston, and M. Young, "Measurement and modeling of depth cue combination: in defense of weak fusion," *Vision Research*, vol. 35, pp. 389–412, 1995.

[9] M. Solh, J. M. Bauza, and G. AlRegib, "3VQM: A vision-based quality measure for DIBR-based 3D videos," in *2011 IEEE International Conference on Multimedia and Expo (ICME)*, Jul. 2011, pp. 1 –6.

[10] Fraunhofer Heinrich Hertz Institut, "H.264/avc software coordination - jm 18.4," .