Time-of-Flight Sensor Fusion with Depth Measurement Reliability Weighting

Sebastian Schwarz, Mårten Sjöström, and Roger Olsson
Mid Sweden University, Sundsvall, Sweden
Email: [sebastian.schwarz, marten.sjostrom, roger.olsson]@miun.se

Abstract—**Accurate scene depth capture is essential for the success of three-dimensional television (3DTV), e.g. for high quality view synthesis in autostereoscopic multiview displays. Unfortunately, scene depth is not easily obtained and often of limited quality. Dedicated Time-of-Flight (ToF) sensors can deliver reliable depth readings where traditional methods, such as stereovision analysis, fail. However, since ToF sensors provide only limited spatial resolution and suffer from sensor noise, sophisticated upsampling methods are sought after. A multitude of ToF solutions have been proposed over the recent years. Most of them achieve ToF super-resolution (TSR) by sensor fusion between ToF and additional sources, e.g. video. We recently proposed a weighted error energy minimization approach for ToF super-resolution, incorporating texture, sensor noise and temporal information. For this article, we take a closer look at the sensor noise weighting related to the Time-of-Flight active brightness signal. We determine a depth measurement reliability function based on optimizing free parameters to test data and verifying it with independent test cases. In the presented double-weighted TSR proposal, depth readings are weighted into the upsampling process with regard to their reliability, removing erroneous influences in the final result. Our evaluations prove the desired effect of depth measurement reliability weighting, decreasing the depth upsampling error by almost 40% in comparison to competing proposals.**

Keywords—*Time-of-Flight, active brightness, sensor-fusion, super-resolution, 3D video, scene depth, depth map upsampling.*

I. INTRODUCTION

The ongoing success of three-dimensional (3D) cinema fuels increasing efforts to spread the commercial success of 3D to new markets. The possibilities of a convincing 3D experience at home, such as three-dimensional television (3DTV), has generated a great deal of interest within the research and standardization community. A central issue for 3DTV is precise scene depth information for Depth Image Based Rendering (DIBR) view synthesis. Acquiring scene depth information is a fundamental task in computer vision, yet complex and error-prone. Dedicated range sensors, such as Time-of-Flight (ToF) sensor, can simplify the scene depth capture process and overcome shortcomings of traditional solutions, like active or passive stereo analysis.

Stereo analysis is a common approach to scene depth extraction. Feature and area analysis between two camera views allows for the reconstruction of depth information based on the camera geometry. However, if parts of the scene are occluded in one view, or if areas have low or repetitive texture, stereo matching produces erroneous results [1]. ToF

sensors can overcome most of these shortcomings. They capture scene depth in real-time, independent of texture structure and occlusions. Admittedly, current ToF sensors deliver only limited spatial resolution and suffer from sensor noise [2]. Thus recent years have shown a multitude of ToF depth upsampling proposals. Many of these utilize a combination of guidance sources in the upsampling process, such as texture information from video [3]–[5], depth measurement reliability [6], [7], temporal filtering [8], or sensor noise information [9], [10]. All aim for the same goal: Fusing low-resolution ToF sensor reading with additional information to achieve ToF super-resolution (TSR).

In 2012, we proposed a TSR approach based on weighted error energy minimization. In our original contribution we enforce spatial similarity between neighboring pixels and assumed that edges in texture correspond to object transitions in depth. To avoid blurring at these important transitions, we weight the spatial smoothness constraints with texture edges, so that pixels at edges are less constrained to be similar. Since texture sources usually contain more edges than actual depth transitions, we mask the texture information with low resolution depth edge information [11]. Recently, we introduced additional weighting functions to address sensor noise and temporal flicker in depth for TSR [12]. For this article, we take a closer look at the ToF sensor noise model [13] and analyze depth measurement reliability weighting based on the received signal intensity, the so called *active brightness*. The main contributions of this publication are the deeper analysis of the active brightness weighting function and the de-coupling of original depth sensor readings from the depth upsampling result by introducing an additional error energy term for the depth measurements.

The remainder of this paper is structured as follows: First we provide a brief explanation of ToF sensors in Sec. II. In Sec. III, we describe our proposed active brightness weighting scheme for TSR. Sec. IV addresses our methodology and Sec. V shows our evaluation results, before we conclude this paper in Sec. VI.

II. TIME-OF-FLIGHT SENSORS

As the name suggests, a Time-of-Flight (ToF) sensor measures the travel time for a reflected light beam from the sender to an object and back to the receiver. There are two categories for ToF sensors [14]: *Pulse Runtime Sensors*, where a pulsed wave is sent out and a clock measures the time that has passed until the reflected signal is received again). And *Continuous Wave Sensors*, where the phase shift between the modulated wave signal and its reflected signal is measured). Continuous

(a) Combined Capture Setup   (b) RGB Texture

(c) ToF Depth   (d) ToF Active Brightness

Fig. 1. Combined ToF+video capture setup (a) and its outputs: RGB Texture (b), low resolution ToF depth (c) and active brightness (d).

wave ToF sensors are predestined for real-time scene depth capture, while Pulse Runtime Sensors suffer from low temporal resolution due to their pulsed nature.

A continuous wave ToF sensor sends out a cosines modulated signal with modulation frequency ν, $s(t) = cos(2\pi\nu t)$. The received signal, $r(t) = A \cdot cos(2\pi\nu t - \Phi)$, consists of the reflected cosines modulated signal with amplitude A, due to signal attenuation, and phase shift Φ, due to reflection. Using the standard equations for light propagation, depth z of an object can be determined by

$$z(\Phi) = \Phi \frac{c}{4\pi\nu}. \qquad (1)$$

However, ToF sensors readings suffer from photon shot noise. Photon shot noise is inversely proportional to the number of collected photons on the sensor [2]. It is the main reason for the limited spatial resolution of ToF cameras, since the single capturing pixel elements on the sensor must be of an adequate size to collect a sufficient number of photons for a reliable depth reading. Frank et al. [13] showed that the effect of photon shot noise can be sufficiently approximated as a zero-mean Gaussian, with standard deviation σ_D, inversely proportional to the active brightness A.

$$\epsilon_D \in N(0, \sigma_D^2) \text{ with } \sigma_D \propto \frac{1}{A} \qquad (2)$$

For 3DTV capture scenarios, ToF sensors are usually coupled with a video camera. Such a ToF+video combination is shown in Fig. 1(a). The RGB camera delivers the necessary video sequence for DIBR view synthesis (Fig. 1(b)), while the ToF sensor delivers depth readings (Fig. 1(c)) and depth measurement reliability in form of the active brightness signal (Fig. 1(d)). Since RGB camera and ToF sensor have differing spatial resolution, the ToF depth requires upsampling for DIBR. The next section explains how to incorporate texture and depth measurement reliability information for ToF super resolution.
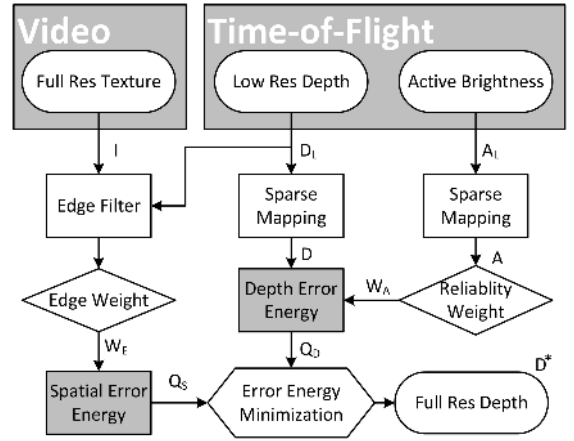

Fig. 2. Proposed ToF upsampling algorithm with texture edge & depth measurement reliability weighting.

III. WEIGHTED OPTIMIZATION TOF UPSAMPLING

Fig. 2 illustrates our proposed algorithm for ToF upsampling in ToF+video combination, with two input sources: The video camera and the ToF sensor. For mathematical context, sources are represented as two-dimensional matrices of pixel values, e.g. texture frame $\mathbf{I} = \{I(x, y); x = 1, ..., X; y = 1, ..., Y\}$ with X and Y as the maximum indices. Mapping low resolution ToF depth $\mathbf{D}_L$ and active brightness $\mathbf{A}_L$ on the corresponding texture frame pixel positions, yields $\mathbf{D}$ and $\mathbf{A}$, matching the video frame target resolution, but with a sparse, irregular value distribution. Time-of-Flight super-resolution $\mathbf{D}^*$ is achieved by minimizing a combination of two energy terms: spatial error energy Q_S and depth error energy Q_D:

$$\mathbf{D}^* = \arg\min_{\hat{\mathbf{D}}}(cQ_S + (1 - c)Q_D), \qquad (3)$$

where $\hat{\mathbf{D}}$ is the argument for the energy minimization and constant c can be used to emphasize one of the input error energies.

The spatial error energy Q_S is formed by the horizontal and vertical similarity errors ϵ_H and ϵ_V, enforcing similarity between neighboring pixel. Edge weighting function W_E uses edge information from texture frame $\mathbf{I}$ and depth map $\mathbf{D}_L$ to relax the spatial similarity constraint at object boundaries, thus creating a piece-wise smooth value distribution for $\mathbf{D}^*$. Creating edge weight $\mathbf{W}_E$ is described in more detail a previous publication [12]. The key concept is to validate texture edge information (edge map $\mathbf{E}_I$ from $\mathbf{I}$) with depth edge information (edge map $\mathbf{E}_D$ from $\mathbf{D}_L$, upsampled to match $\mathbf{E}_I$) to mask out unnecessary edges in the edge weighting map $\mathbf{W}_E$. We call this approach single-weighted ToF super-resolution (S-TSR).

$$\epsilon_H(x, y) = \hat{D}_{(x,y)} - \hat{D}_{(x+1,y)}, \qquad (4)$$

$$\epsilon_V(x, y) = \hat{D}_{(x,y)} - \hat{D}_{(x,y+1)}, \qquad (5)$$

$$Q_S = \sum_{x,y=1}^{X,Y} W_{E(x,y)}\epsilon_H^2(x, y) + \sum_{x,y=1}^{X,Y} W_{E(x,y)}\epsilon_V^2(x, y), \qquad (6)$$

$$W_E(x, y) = 1 - E_I(x, y) \cdot E_D(x, y) \qquad (7)$$

In the original publication [11], the spatial error energy Q_S was used directly to fill the missing values in $\mathbf{D}$, thus adopt-

ing erroneous depth readings directly into the ToF super-resolution result. For this paper, erroneous depth measurements are suppressed by introducing the depth error energy Q_D. Depth error ϵ_D enforces similarity between the sparse ToF depth readings in $\mathbf{D}$ and the final super-resolution result $\mathbf{D}^*$. Transferred into energy terms, we derive depth error energy Q_D. Depth measurement reliability weight $\mathbf{W}_D$ can be used to remove erroneous depth readings, suppress low reliability, and emphasize high reliability depth readings in the upsampling process. Thus, we call this approach double-weighted ToF super-resolution (D-TSR).

$$\epsilon_D(x,y) = D(x,y) - \hat{D}(x,y) \tag{8}$$

$$Q_D = \sum_{x,y=1}^{X,Y} W_D(x,y)\epsilon_D^2(x,y) \tag{9}$$

The depth measurement reliability weight $\mathbf{W}_D$ is directly derived from the sparse active brightness map $\mathbf{A}$ with

$$W_D(x,y) = \begin{cases} \left(\frac{A(x,y)}{l_l}\right)^\alpha, & l_l < A(x,y) < l_u \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

The concept behind this weighting function is derived from the photon shot noise model in Eq. 2, together with practical observations on ToF sensors [15]. Exponent α defines the basic weighting function based on active brightness $\mathbf{A}$. The range $[l_l; l_u]$ limits the active brightness within which ToF depth readings are considered useful. Eq. 2 supports the assumption that the best results are achieved with the highest active brightness. Unfortunately this is not completely true. At high active brightness levels, photo-generated electrons flood the capturing pixel element, causing erroneous depth readings. Therefore it is beneficial to define an active brightness value range within which the sensor is sufficiently saturated. Outside this range, the depth readings should be removed from our determined system. $\mathbf{W}_D$ is then normalized to form a weighting value range of $[0; 1]$, with 1 for the most reliable and 0 for discarded depth readings. Based on this weighting, depth measurements with high reliability have a stronger influence on the upsampling process, while values with lower reliability have less.

The basic concept of D-TSR was introduced in the above mentioned previous publication [12]. The main contributions for the current article are the de-coupling between sensor readings and depth upsampling result through the weighted depth error energy Q_D, and the following deeper analysis of the depth measurement reliability weight $\mathbf{W}_D$.

IV. METHODOLOGY

A detailed discussion of the different factors influencing the spatial error energy Q_S is provided in [11]. Within this paper, we analyze the presented depth measurement reliability function W_D to determine an appropriate active brightness weighting function, and the relationship between spatial error energy Q_S and depth error energy Q_D for best ToF upsampling results, in a mean squares error sense. In order to determine good choices for the addition of depth measurement reliability, we define our ToF upsampling approach as system T with inputs $\mathbf{I}$, $\mathbf{D}_L$ and $\mathbf{A}_L$, output $\mathbf{D}^*$, and four variables: Exponent α, active brightness limits l_l and l_u from Eq. 10, and constant c from Eq. 3:

$$\mathbf{D}^* = T(\mathbf{I}, \mathbf{D}_L, \mathbf{A}_L, [\alpha, l_l, l_u, c]) \tag{11}$$

To rate the performance of system T, we created a set of 12 different test cases. Each test case consists of four single frames: Low resolution depth and active brightness from a Fotonic B-70 ToF sensor, plus a target resolution texture frame for the upsampling process and a depth upsampling reference. The depth reference is created by averaging the ToF depth readings over 250 frames (10s), thus reducing the photon shot noise from Eq. 2, and followed by an upsampling by a factor of 8 using S-TSR. As rating metric we choose the mean squared error (MSE) between system output $\mathbf{D}^*$ and the reference depth. Lower MSE equals to better system performance. The test cases cover several scenarios (checkerboard, reflectivity gradient, different still scenes) with varying camera orientation and active brightness saturation. Minimizing MSE over all 12 test cases yields a set of optimized parameters for the upsampling process: $\alpha = 0.185$, $l_l = 100$, $l_u = 4000$ and $c = 0.52$.

With these results we evaluate our approach against the previous S-TSR implementation and Joint-Bilateral upsampling (JBU, [3]). This time with six test cases which have not been part of the previous parameter estimation. The individual test cases are: Checkerboard filling the complete image, with well-exposed (case 1) and slightly under-exposed (case 2) active brightness. Gray-scale gradient from white to black, filling the complete image, with well (case 3) and slightly over-exposed (case 4) active brightness. Two different still color scenes with maximum depth of 3m (case 5) and 7m (case 6). While cases 1 to 4 represent lab conditions, cases 5 and 6 represent real-life applications. The evaluation metric is MSE. JBU is a very popular texture guided depth upsampling approach. It was first proposed in 2007 and might not be state-of-the-art. However, its easy implementation leads to a widespread acceptance and it is often presented in ToF upsampling publications. Therefore JBU is considered a good choice of reference to allow comparisons between different proposals using different test data.

V. RESULTS

Fig. 3 shows the MSE comparison between different depth upsampling approaches, normalized with respect to the JBU upsampling. D-TSR outperforms the competing approaches for every case, with an overall decrease in MSE of 37.5%, compared to JBU, respectively 24% compared to S-TSR. These results support the validity of depth measurement reliability weighting for ToF upsampling. The positive influence of the depth measurement reliability weight is especially noticeable between the cases 1 & 2 and 3 & 4. Both case combinations each represent identical content, but at different active brightness saturation. With the former (case 1 and 3) well saturated in terms of active brightness, and the latter slightly under-(case 2) or over-exposed (case 4). The visual results are similar for all cases. Within the extent of this paper, Fig.4 presents only images for case 6. Please note the relationship between high noise in original depth (a) and low active brightness (b), especially in the background on the right. This noise is leveled out in the reference depth (c), but still visible in the
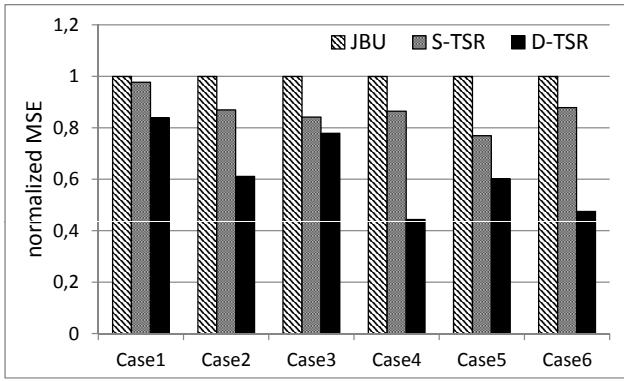
Fig. 3. Depth upsampling error in MSE for different upsampling approaches by a factor of 8. Normalized with respect to JBU results.

(a) Original Depth   (b) Active Brightness

(c) Reference Depth   (d) JBU
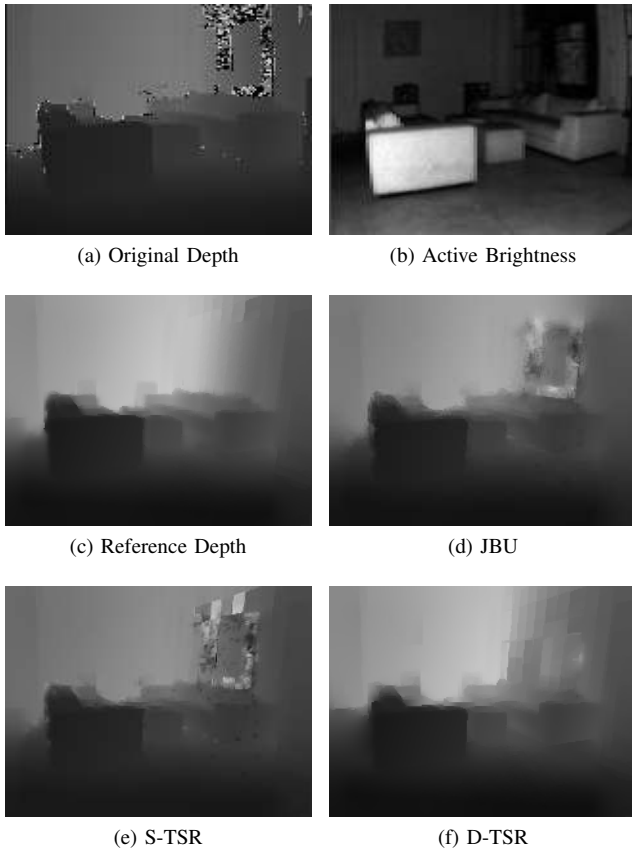
(e) S-TSR   (f) D-TSR

Fig. 4. Test case 6: Original depth (a), active brightness (b), reference depth (c) and upsampling results for JBU (d), S-TSR (e) and proposed D-TSR (f).

JBU (d) and S-TSR (e) upsampling result. The proposed D-TSR upsampling (f) delivers a result close to the reference depth. For a detailed evaluation of S-TSR with respect to JBU and further competing depth upsampling approaches, please be referred to the previous publications [11], [12].

VI. CONCLUSIONS

In this paper we have presented an important addition to our recently proposed TSR approach based on error energy minimization [12]. Original ToF sensor readings are weighted into the upsampling process with regard to their reliability based on the active brightness. We determined an eligible depth measurement reliability weighting function based on a parameter estimation process. Our results show a significant decrease in depth upsampling error compared to previous proposals, when validated on independent data. Please be aware that one should not generalize the specific weighting function variables $[\alpha, l_l, l_u, c]$. The values presented in this paper only apply to the Fotonic B-70 ToF sensor. Other ToF sensors might require an adaption of these system variables. However the overall statement for this contribution remains intact: Depth measurement reliability weighting significantly increases the quality for ToF super-resolution, thus simplifying scene depth acquisition and improving 3D video content quality.

REFERENCES

[1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, no. 1-3, pp. 7–42, 2002.

[2] R. Lange and P. Seitz, "Solid-state time-of-flight range camera," *Quantum Electronics, IEEE Journal of*, vol. 37, no. 3, pp. 390–397, 2001.

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

[4] 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), MIT Press, 2005.

[5] D. Min, J. Lu, and M. Do, "Depth video enhancement based on weighted mode filtering," *Image Processing, IEEE Trans. on*, vol. 21, no. 3, pp. 1176–1190, 2012.

[6] 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.

[7] M. Reynolds, J. Dobos, L. Peel, T. Weyrich, and G. Brostow, "Capturing time-of-flight data with confidence," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 2011.

[8] J. Zhu, L. Wang, J. Gao, and R. Yang, "Spatial-temporal fusion for high accuracy depth maps using dynamic mrfs," *Pattern Analysis and Machine Intelligence, IEEE Trans. on*, vol. 32, pp. 899–909, 2010.

[9] 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*, 2008.

[10] B. Huhle, T. Schairer, P. Jenke, and W. Straſer, "Fusion of range and color images for denoising and resolution enhancement with a non-local filter," *Comput. Vis. Image Underst.*, vol. 114, no. 12, 2010.

[11] S. Schwarz, M. Sjöström, and R. Olsson, "Depth map upscaling through edge weighted optimization," in *Proceedings of the SPIE, vol 8290: Three-Dimensional Image Processing (3DIP) and Applications II*, 2012.

[12] S. Schwarz, M. Sjöström, and R. Olsson, "A weighted optimization approach to time-of-flight sensor fusion," *IEEE Transactions on Image Processing*, vol. 23(1), pp. 214–225, 2014.

[13] M. Frank, M. Plaue, H. Rapp, U. Köthe, B. Jähne, and F. A. Hamprecht, "Theoretical and experimental error analysis of continuous-wave time-of-flight range cameras," *Optical Engineering*, vol. 48, no. 1, 2009.

[14] T. Kahlmann and H. Ingensand, "Calibration and improvements of the high-resolution range-imaging camera swissranger," in *Proceedings of the SPIE, vol 5665: Videometrics VIII*, 2006.

[15] B. Büttgen, T. Oggier, M. Lehmann, R. Kaufmann, and F. Lustenberger, "CCD/CMOS lock-in pixel for range imaging : Challenges, limitations and state-of-the-art," *Measurement*, vol. 103, 2005.