RELIABLE FUSION OF TOF AND STEREO DEPTH DRIVEN BY CONFIDENCE MEASURES

Giulio Marin, Pietro Zanuttigh - University of Padova, Italy - {maringiu, zanuttigh}@dei.unipd.it
Stefano Mattoccia - University of Bologna, Italy - stefano.mattoccia@unibo.it

Introduction

Most common technologies for depth estimation:

- **Passive stereo vision systems**: require a pair of standard cameras and can provide a high resolution depth estimation in real-time. However, results are not completely reliable and strongly depend on scene characteristics.
- **Time-of-Flight (ToF) cameras**: are able to robustly estimate in real time the 3D geometry of a scene but they are limited by a low spatial resolution and a high level of noise in their measurements.

Proposed method

The proposed algorithm is divided into three steps:

1. Estimation of high resolution depth from ToF camera with associated confidence measure.
2. Estimation of depth map from stereo cameras and associated confidence measure.
3. Fusion of depth maps using the confidence measures enforcing local consistency.

ToF depth map and confidence

Depth map
ToF data are projected on the left color camera and then interpolated combining cross bilateral filtering and color segmentation.

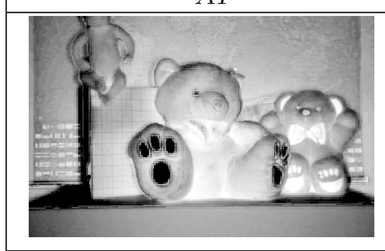
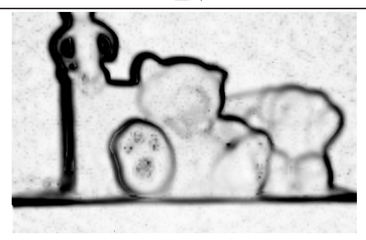
Confidence

- P_{AI}: amplitude and intensity of ToF signal.

$$\sigma_z = \frac{c}{4\pi f_{mod}} \frac{\sqrt{I/2}}{A} \quad \Rightarrow \quad \sigma_d = bf \, \frac{\sigma_z}{z^2 - \sigma_z^2} \tag{1}$$

- P_{LV}: accounts for local depth variance.

$$D_l^{TOF} = \frac{1}{\left|\mathcal{N}(p_i^{TOF})\right|} \sum_{j \in \mathcal{N}(p_i^{TOF})} |z_i - z_j| \tag{2}$$

P_{AI}	P_{LV}	$P_T = P_{AI} P_{LV}$
		

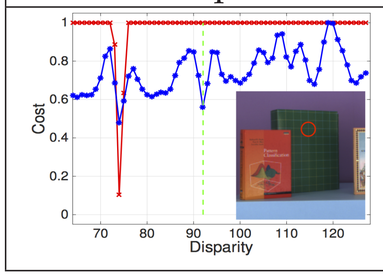
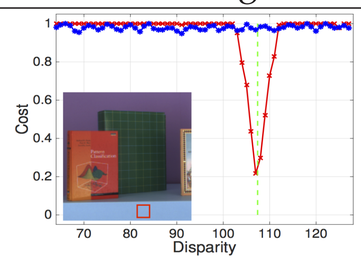
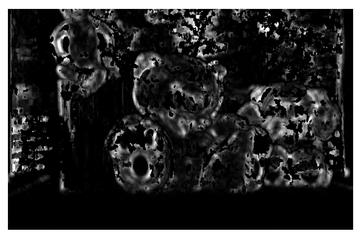
Stereo depth map and confidence

Depth map
A high resolution disparity map is inferred by global or semi-global stereo vision algorithms.

Confidence

$$P_S = \frac{\Delta C^l}{C_1^l} \left(1 - \frac{\min\{\Delta d^l, \gamma\}}{\gamma}\right) \left(1 - \frac{\min\{\Delta d^{lg}, \gamma\}}{\gamma}\right) \tag{3}$$

- $C_1 = C(d_1) = \min_d C(d), \quad C_2$ s.t. $|d_2 - d_1| \leq 1$
- l: local, g: global
- $\Delta C^l = C_2^l - C_1^l, \quad \Delta d^l = |d_2^l - d_1^l|, \quad \Delta d^{lg} = |d_1^l - d_1^g|$

Periodic pattern	Uniform region	P_S
		

Local cost, Global cost, Ground truth

Data fusion

$$\Omega'_f(d) = \sum_{g \in \mathcal{A}} \left(P_T(g) \mathcal{P}_{f,g,T}(d) \; + \; P_S(g) \mathcal{P}_{f,g,S}(d) \right) \tag{4}$$

- $P_T(g)$ and $P_S(g)$: confidence maps.
- $\mathcal{P}_{f,g,T}(d)$, $\mathcal{P}_{f,g,S}(d)$: plausibility terms.

$$\mathcal{P}_{f,g}(d) = e^{-\frac{\Delta_{f,g}}{\gamma_s}} \cdot e^{-\frac{\Delta^{\psi}_{f,g}}{\gamma_c}} \cdot e^{-\frac{\Delta^{\psi}_{f',g'}}{\gamma_c}} \cdot e^{-\frac{\Delta^{\omega}_{g,g'}}{\gamma_t}} \tag{5}$$

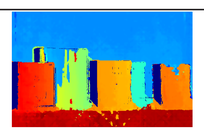
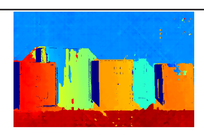
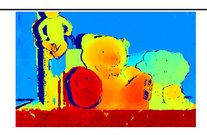
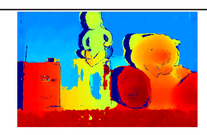
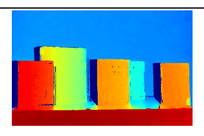
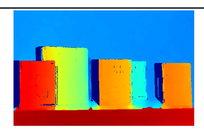
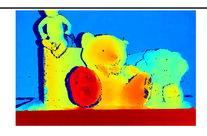
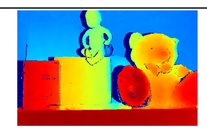
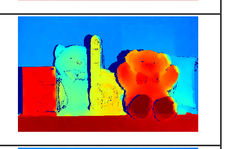
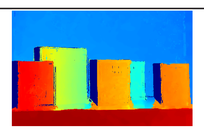
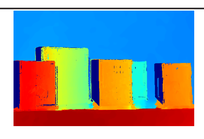
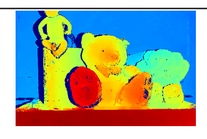
- f, g and f', g': points in the left and right image
- Δ: spatial proximity; Δ^{ψ}, Δ^{ω}: color similarity

Comparison

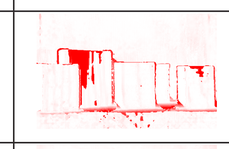
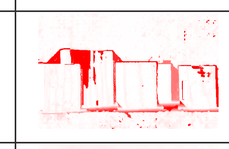
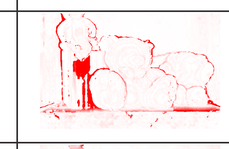
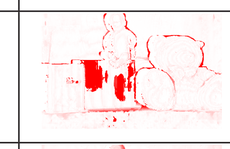
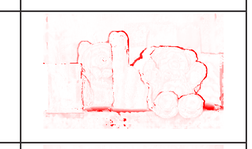
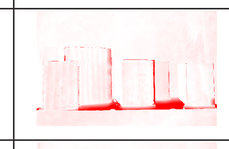
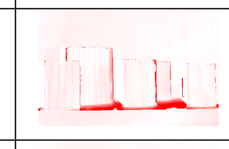
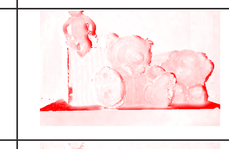
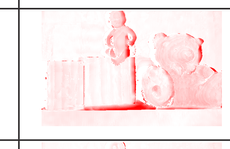
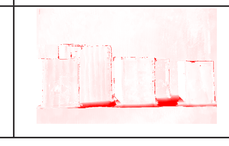
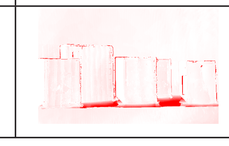
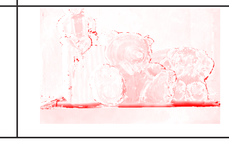
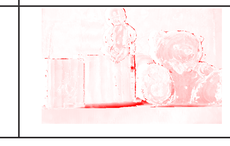
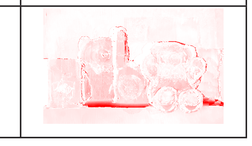
Scene	1	2	3	4	5	Avg.
ToF Int.	9.83	10.33	14.43	8.68	15.12	11.67
Stereo	19.17	27.83	18.06	25.52	11.49	20.42
Fusion	**7.40**	9.33	**6.92**	**6.30**	**8.39**	**7.67**
[1]	7.43	**9.27**	12.60	7.99	13.01	10.06
[2]	8.49	9.92	11.44	9.88	15.19	10.98
[3]	9.04	10.04	13.04	9.52	14.03	11.13
[4]	10.98	13.19	9.83	13.93	13.10	12.21
Ideal	2.50	2.60	3.22	2.42	3.16	2.78

Experimental results

Depth maps

	1	2	3	4	5
ToF					
Stereo					
Fusion					

Errors

	1	2	3	4	5
ToF					
Stereo					
Fusion					

[1] Dal Mutto, C., Zanuttigh, P., Mattoccia, S., Cortelazzo, G.M.: Locally consistent tof and stereo data fusion. In: Workshop on Consumer Depth Cameras for Computer Vision (ECCV Workshop), 2012.
[2] Yang, Q., Yang, R., Davis, J., Nister, D.: Spatial-depth super resolution for range images. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007.
[3] Zhu, J., Wang, L., Yang, R., Davis, J.: Fusion of time-of-flight depth and stereo for high accuracy depth maps. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008.
[4] Dal Mutto, C., Zanuttigh, P., Cortelazzo, G.: Probabilistic tof and stereo data fusion based on mixed pixels measurement models. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2015.

http://lttm.dei.unipd.it/paper_data/eccv16