Real-time 3D Imaging System Operating in Low-Sensing Mode

M. Georgiev, A. Gotchev
Tampere University of Technology,Finland
mihail.georgiev@tut.fi, atanas.gotchev@tut.fi

I. Introduction

Three-dimensional (3D) imaging refers to a set of techniques aimed at sensing 3D visual scenes, their processing and recreation on various displays. A common approach in sensing is to employ multi-sensor hardware composed of one or more color (RGB) cameras and a depth sensor. The non-confocal setting of the two devices requires a post-processing stage of reprojection, resampling and data fusion to bring both, color and depth, modalities into common grid and viewing perspective.

We have developed an end-to-end 3D imaging system, where the sensing end consists of high-definition (HD) RGB camera and a Time-of-flight (ToF) depth sensor. The system aims at real-time performance on power-constrained hardware and sensing conditions that we characterize as 'low-sensing environment' (LSE). LSE can be caused by low-light sensing conditions, enforcement of low-power hardware, and small pixel resolution. Addressing such cases is important given the aim for reducing the size and power of ToF devices in an attempt to integrate them in portable multi-functional devices meant to operate also in low-light conditions. While in such cases the data appears extremely noisy, we demonstrate that it can be effectively post-processed in order to reach the same performance of 3D view-synthesis applications as if the device was working in normal operating mode.

In this demo, we present a practical implementation of the system and more specifically its denoising module, working in real time. The latter is composed by a module suppressing a spatially-correlated sensor effect known as *fixed-pattern noise* (FPN) and a module implementing a novel non-local method for ToF data denoising [1]. FPI is characteristic for ToF sensors working in LSE and has to be handled first [3]. The remaining noise propagates to the harmonic components of the measured signal and has to be handled with care. We have devised an effective denoising method, which favors the use of a complex-valued representation of the ToF sensed signal and makes use of its naturally stabilized noise variance [1].

II. System Modules

Our system shares the same structure as proposed in [2] and illustrated in Fig. 1. It includes modules for ToF data de-noising, projection alignment, non-uniform image up-sampling, dis-/occlusion detection and inpainting, and virtual-view synthesis for 3D data manipulation and visualization.

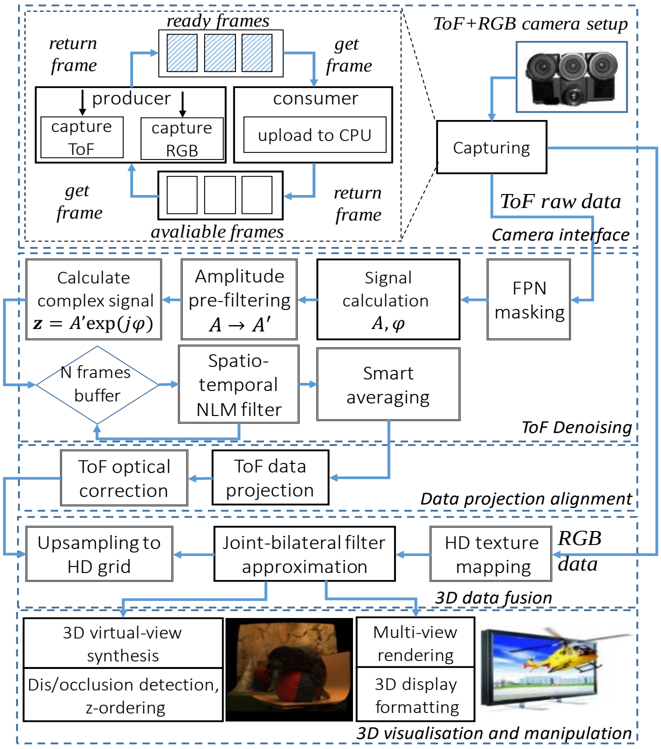

Fig 1. System modules of the proposed 3D imaging system realization.

A. Multi-sensor data capture module

Simultaneous multi-sensor capture is a challenging problem since different sensors employ different sensing principles. We have proposed a software triggering mode that minimizes possible lag and thus results in near-synchronous multi-sensor data acquisition [2] with no observable delay. In our solution, we create a separate capturing thread to trigger all sensors, whenever the overall setup gets ready to update device memory and then frames are obtained in a *producer-consumer* manner (c.f. Fig.1).

B. Denoising module

The LSE regime of the system has been pushed to limits by reducing both the power consumption and setting the ToF sensor to operate with very short integration time (e.g. 50 μs instead of 2 ms). The system operates in a mode where the sensitivity is decreased at least 100 times, which calls for implementing our dedicated ToF denoising module [1]. The module consists of two stages, first for FPN removal and second for denoising the amplitude and phase of the sensed signal.

The FPN is ubiquitous for the current sensor technology (e.g. CMOS or CCD) [3]. While it can be neglected for normal operating mode, its appearance becomes dominant in LSE, since its levels start to be comparable to the sensor charge intensities [3, 1]. FPN removal is mandatory before applying the denoising step of the sensed signal. The realization of FPN suppression stage considers FPN as a pseudo-periodic pattern which can be filtered by an adaptive FIR notch filter [3].

The denoising stage implementation is based on an iterative complex-valued Non-local Means (NLM) filter [1], where the patch similarity is searched in *spatio-temporal domain* (e.g. video buffer of 3 frames). Flickering and motion artifacts are tackled by applying *smart-averaging* as proposed in [4], employing an adaptive threshold driven by the expected noise variance of the complex signal. The computational overhead has been addressed by approximating NLM using summed-area tables (SATs), and employing low-complexity fixed-point arithmetic [5], SIMD vectorization assembly, and multi-core parallelization.

C. 3D data fusion

The 3D fusion stage implements an iterative non-uniform resampling procedure of Richardson type [6]. More specifically, the ToF samples are iteratively projected on the HD grid of the RGB camera and then regularized by a cross-modality bilateral filter.

D. Virtual-view synthesis and multi-view display visualization

The output of the 3D imaging system is demonstrated for two 3D applications – virtual-camera view synthesis and auto-stereoscopic 3D display visualization. The latter requires synthesis of a number of perspective views. While the virtual-view synthesis is handled by ordinary surface-mesh rendering approach [2], the multi-view synthesis case is tackled by *backward depth projection layering* in disparity domain [7]. The reference depth map is partitioned on a segment-by-segment basis, producing a complete virtual view layer map. This approximation facilitates the use of connectivity information for segment-based forward warping of the reference layer map. The warped layer map is used to guide the dis-occlusions *in-painting process* of the synthesized texture map [8].

III. Software Realization

The denoising performance of the proposed denoising module is illustrated in Fig. 2. The quality improvement due to the fast-approximation scheme of spatio-temporal NLM is seen in Fig. 2 (a-c) (3 frames). The compensation of motion artifacts by smart-averaging filter is illustrated further in Fig 3(d), where much crispier scene edge and fine details are obtained. The denoising and dis-occlusion detection performance can be also visually assessed for the virtual-view synthesis output (c.f. Fig. 3 (e-f)).

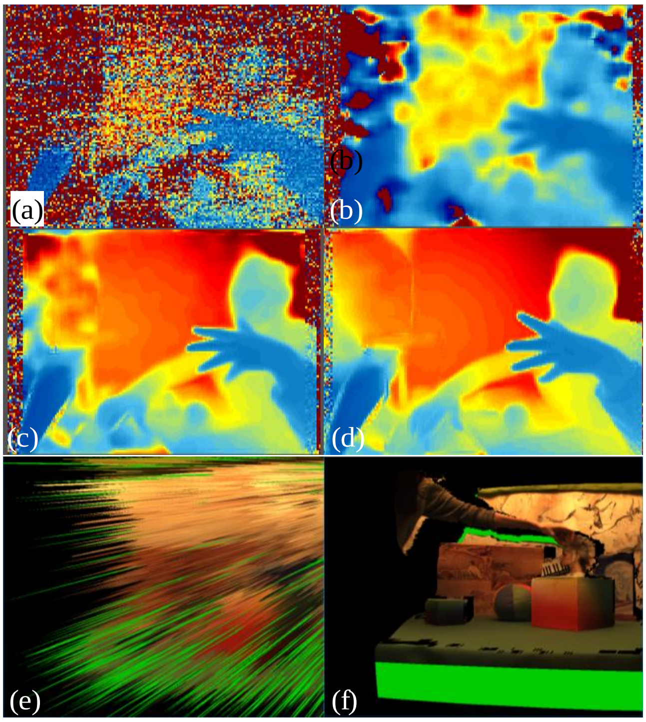

Fig 2. Exemplified denoising performance of the 3D imaging system: a) unprocessed LSE noisy input, denoised by b) complex domain NLM filter, c) spatio-temporal approximated scheme (3 frames), d) smart-averaging applied; virtual-view synthesis demonstrated for e) LSE noisy input, d) denoising applied. (dis-/occlusion areas are painted in green and black).

IV. References

[1] M. Georgiev, R. Bregovic and A. Gotchev, "Time-of-flight range measurement in low-sensing environment: Noise analysis and complex-domain non-local denoising," *IEEE Trans. on Image Processing*, vol. 27, no. 6, pp. 2911-2926, 2018.

[2] A. Chuchvara, M. Georgiev and A. Gotchev, "A speed-optimized RGB-Z capture system with improved denoising capabilities," in *Proc. of SPIE, Image Processing: Algorithm and Systems*, vol. 9019, no. F, pp. 1533-1537, San Francisco, CA, Feb. 2014.

[3] M. Georgiev, R. Bregovic and A. Gotchev, "Fixed-pattern noise modelling and removal in Time-of-flight sensing," *IEEE Trans. on Instr. and Measurement*, vol. 65, no. 4, pp. 808-820, April 2016.

[4] A. Chuchvara, M. Georgiev and A. Gotchev, "A framework for fast low-power multi-sensor 3D scene capture and reconstruction," in *Int. Workshop on Biometric Authentication*, vol. 8853, pp. 40-53, Sofia, 2014.

[5] M. Georgiev, A. Gotchev and M. Hannuksela, "Real-time denoising of ToF measurements by spatio-temporal non-local mean filtering," in *Proc, of IEEE Int. Conf. on Multim. and Expo Workshops*, p.6, San Jose, CA, 2013.

[6] M. Georgiev, A. Gotchev and M. Hannuksela, "Joint de-noising and fusion of 2D video and depth map sequences," in *IEEE Int. Conf, on Multimedia and Expo Workshops*, p. 4, San Jose, CA, 2013.

[7] A. Chuchvara, M. Georgiev and A. Gotchev, "CPU-efficient free view synthesis based on depth layering," in *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video*, p. 4, Budapest, 2014.

[8] A. Chuchvara, O. Suominen, M. Georgiev and A. Gotchev, "Speed-optimized free-viewpoint rendering based on depth layering," in *Proc. of IEEE 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video*, p.4, Lisbon, July 2015.