Image and Vision Computing 41 (2015) 26–41

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Editor's Choice Article

Unified multi-lateral filter for real-time depth map enhancement☆

Frederic Garcia [a,*], Djamila Aouada [b], Bruno Mirbach [a], Thomas Solignac [a], Björn Ottersten [b]

[a] PTU Optical, IEE S.A., Luxembourg
[b] Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, Luxembourg

ARTICLE INFO

Article history:
Received 25 June 2014
Received in revised form 25 April 2015
Accepted 24 June 2015
Available online 3 July 2015

Keywords:
Depth enhancement
Data fusion
Sensor fusion
Multi-modal sensors
Adaptive filters
Active sensing
Time-of-Flight

ABSTRACT

This paper proposes a unified multi-lateral filter to efficiently increase the spatial resolution of low-resolution and noisy depth maps in real-time. Time-of-Flight (ToF) cameras have become a very promising alternative to stereo-based range sensing systems as they provide depth measurements at a high frame rate. However, there are actually two main drawbacks that restrict their use in a wide range of applications; namely, their fairly low spatial resolution as well as the amount of noise within the depth estimation. In order to address these drawbacks, we propose a new approach based on sensor fusion. That is, we couple a ToF camera of low-resolution with a 2-D camera of higher resolution to which the low-resolution depth map will be efficiently upsampled. In this paper, we first review the existing depth map enhancement approaches based on sensor fusion and discuss their limitations. We then propose a unified multi-lateral filter that accounts for the inaccuracy of depth edges position due to the low-resolution ToF depth maps. By doing so, unwanted artefacts such as texture copying and edge blurring are almost entirely eliminated. Moreover, the proposed filter is configurable to behave as most of the alternative depth enhancement approaches. Using a convolution-based formulation and data quantization and downsampling, the described filter has been effectively and efficiently implemented for dynamic scenes in real-time applications. The experimental results show a sensitive qualitative as well as quantitative improvement on raw depth maps, outperforming state-of-the-art multi-lateral filters.

1. Introduction

Time-of-Flight cameras have become an alternative 3-D sensing system to stereo-based range sensing systems such as stereo vision systems, laser scanners or structured light. They present several advantages such as simultaneously providing intensity and depth information for every pixel at a high frame rate. Moreover, the recent advances in industrializing and producing economic, compact, robust to illumination changes and low-weight ToF cameras is starting to have an impact on commercial applications [1–3]. This is the case of the Xbox One's Kinect version 2, which uses a ToF based depth camera. It has shown a significant improvement over the structured light based Kinect version that was shipped with the Xbox 360 [3]. However, as a compromise for their higher robustness to ambient conditions; i.e., larger working temperature range and higher reliability under sun light; and in contrast to depth sensing systems intended for gaming applications or research purposes, the resolution of ToF cameras that are used in automotive or industrial automation is still far below the resolution of common 2-D cameras. Indeed, a very few ToF cameras slightly exceed the Quarter-Quarter Video Graphics Array (QQVGA) resolution, i.e., (160×120) pixels. Note that consumer depth cameras such as Microsoft Kinect or Asus Xtion Pro Live are not compliant to the standards in automotive or industrial automation. Hence, their rather low spatial resolution and the high noise level within their depth measurements are currently restricting their use.

A suitable solution to handle the limited resolution of ToF cameras is sensor fusion [4–11,47]. That is, depth data is complemented with data from a coupled sensor, usually 2-D cameras [12]. Previous works in ToF and 2-D data fusion have proven to deliver dense depth maps at near real-time frame rates, outperforming, in some cases, alternative 3-D sensing systems [4,6]. In this paper, we first introduce existing low-resolution depth map enhancement methods and discuss their limitations. Then, we propose a multi-lateral filter that efficiently tackles inaccurate depth measurements from low-resolution depth maps, producing an enhanced depth map of the same resolution as the considered 2-D guidance image. To this end, we rely on a confidence measure, to be incorporated in the filter formulation, that describes the reliability of each depth measurement. In addition, depth pixels with low confidence will be further treated to overcome unwanted artefacts resulting from data fusion. We also discuss the practical issues such as the configuration of the filter's parameters in order to make it behave as most of the existing multi-lateral filters in the literature. That is, we propose a unified formulation for fusion based depth enhancement filters. Finally,

☆ Editor's Choice Articles are invited and handled by a select rotating 12 member Editorial Board committee. This paper has been recommended for acceptance by Philippos Mordohai.

* Corresponding author at: IEE S.A. (PTU Optical) ZAE Weiergewan 11, rue Edmond Reuter L-5326 Contern Luxembourg.
E-mail addresses: frederic.garcia@iee.lu (F. Garcia), djamila.aouada@uni.lu (D. Aouada), bruno.mirbach@iee.lu (B. Mirbach), thomas.solignac@iee.lu (T. Solignac), bjorn.ottersten@uni.lu (B. Ottersten).

http://dx.doi.org/10.1016/j.imavis.2015.06.008
0262-8856/

we propose a mathematical formulation that enables the filter to be used in real-time applications. We remark that the proposed filter was already introduced in [13] where we provided advice on how to speed up its performance as well as some preliminary results. Herein, we provide a detailed description of its components as well as a comprehensive qualitative and quantitative evaluation and comparison against alternative depth enhancement filters. In addition, we discuss its possible configurations to be adapted for different sensing system technologies and/or applications.

The remainder of the paper is organized as follows: Section 2 covers the background and a literature review on depth map enhancement by means of low-level data fusion. In Section 3 we present our multi-lateral filter as well as the confidence measure that is used to combine depth and 2-D information. Section 4 proposes a mathematical formulation that enables for a real-time implementation. In Section 5, we compare and quantify the proposed filter with alternative depth enhancement methods using synthetic scenes from the Middlebury dataset as well as our own recorded sequences. Finally, concluding remarks are provided in Section 6.

2. Background and related work

The idea of coupling a ToF camera and a 2-D video camera in a hybrid multi-camera rig is a convenient strategy to overcome the limitations of ToF cameras. Indeed, sensor fusion exploits the advantages of each of the cameras in the camera rig avoiding their individual drawbacks. In this paper we target a low-level data fusion in contrast to higher fusion levels that concern post-processed data [14]. To this end, we need first to ensure a good data matching between the cameras that constitute the camera rig. Existing 3-D warping techniques, i.e., forward warping techniques [15,16,11], assign to each depth pixel its corresponding RGB pixel. However, low-level data fusion resulting in a high-resolution depth map requires the opposite mapping, that is, to assign to each high-resolution 2-D pixel its corresponding depth pixel. Note that this so-called backward warping is far from a trivial task due to the need of determining the distance-dependent disparity between each 2-D and depth pixel correspondences. Herein, we use the real-time backward warping algorithm proposed in [17], an iterative approach that addresses backward warping mapping with an accuracy of half the high-resolution 2-D pixel.

Among the early results on low-resolution depth map enhancement, we emphasize those that use Markov Random Fields (MRFs) to fuse different sources of data. Diebel et al. [5] proposed a direct application of MRFs to range sensing with very promising results. Their approach was later extended by Gloud et al. [7] in the field of robotics, enabling a robot to detect and recognize objects from enhanced depth data. However, the authors did not tackle depth discontinuities along depth edges in a dedicated. This leads to strange artefacts within depth edges. Besides, the evaluation of depth enhancement methods based upon an MRF is in general computationally intensive and thus not suitable if real-time is required. Park et al. [18] proposed a dedicated edge weighting scheme to be used in combination with a non-local means regularization in a least-square optimization approach. User interaction and depth edge mark-up might be required to obtain optimal results as well as image segmentation to define the colour similarity for their edge confidence weighting. Yang et al. [9] proposed to improve the quality of a given low-resolution depth map through an iterative refinement module based upon a cost volume. But again, the authors neither tackle depth discontinuities and real-time remained an open question. An alternative solution for depth enhancement is to consider bilateral filtering [19,20] while fusing ToF and 2-D data. Although bilateral filtering is commonly used in image processing and/or computer vision applications for 2-D noise removal and edge-preserving, it has become the basis of recent depth enhancement approaches, outperforming, in some cases, those approaches based upon an MRF or iterative techniques. A more noticeable aspect is the possibility of performing in real-time thanks to its latest implementation strategies [21–23]. Indeed, the real-time aspect was one of the strongest reasons that motivated us to opt for bilateral filtering when fusing the data. In the following, we introduce the bilateral filter. Later, we review the literature on depth enhancement approaches based on bilateral filtering.

2.1. Bilateral filter

The bilateral filter was first introduced by Tomasi et al. as an alternative to iterative approaches for image noise removal [20] such as anisotropic diffusion, weighted least squares, and robust estimation [19]. This non-iterative filter formulation is a simple weighted average of the local neighbourhood samples, i.e.,

$$\mathrm{J}_{\mathrm{BF}}(\mathbf{p}) = \frac{\sum_{\mathbf{q}\in N(\mathbf{p})} w_{\mathrm{BF}}(\mathbf{p},\mathbf{q}) \cdot \mathrm{I}(\mathbf{q})}{\sum_{\mathbf{q}\in N(\mathbf{p})} w_{\mathrm{BF}}(\mathbf{p},\mathbf{q})}, \tag{1}$$

where $N(\mathbf{p})$ is the neighbourhood at the pixel indexed by the position vector $\mathbf{p} = (i,j)^T$, with i and j indicating the row, respectively column corresponding to the pixel position. The weights are computed based on spatial and radiometric distances between the centre of the considered sample and the neighbouring samples. That is, the kernel is decomposed into a spatial weighting term $f_S(\cdot)$ that applies to the pixel position $\mathbf{p}$, and a range weighting term $f_I(\cdot)$ that applies to the pixel value $\mathrm{I}(\mathbf{p})$ as follows

$$w_{\mathrm{BF}}(\mathbf{p},\mathbf{q}) = f_S(\mathbf{p},\mathbf{q}) \cdot f_I(\mathrm{I}(\mathbf{p}),\mathrm{I}(\mathbf{q})). \tag{2}$$

The weighting terms $f_S(\cdot)$ and $f_I(\cdot)$ are generally chosen to be Gaussian functions with standard deviations σ_S and σ_I, respectively. The parameter σ_S defines the spatial dimension of the kernel, i.e., the size of the pixel neighbourhood $N(\mathbf{p})$, whereas σ_I defines the resolution in I for edge preserving. That is, pixels $\mathbf{p}$ and $\mathbf{q}$ influence each other as long as their values differ within the range $\pm\, \sigma_I$. When both standard deviations are well chosen, the resulting filtered image J_{BF} is a smoothed version of I, with preserved edges and significantly reduced noise level. Thus, by replacing I by a low-resolution depth map D, the resulting filtered depth map J_{BF} will present much less noise even though the low-resolution problem will not be tackled.

2.2. Joint Bilateral Upsampling (JBU)

In order to tackle the low-resolution problem, most of the recent depth enhancement techniques by fusion [4,24,6,13] turn to the Joint Bilateral Upsampling (JBU) filter, introduced by Kopf et al. in [8]. The JBU filter is a modification of the bilateral filter in which the data to be filtered and the data through which the weighting term is computed are not coincident. That is, I in (1) is replaced by the low-resolution depth map D, i.e.,

$$\mathrm{J}_{\mathrm{JBU}}(\mathbf{p}) = \frac{\sum_{\mathbf{q}\in N(\mathbf{p})} w_{\mathrm{BF}}(\mathbf{p},\mathbf{q}) \cdot \mathrm{D}(\mathbf{q})}{\sum_{\mathbf{q}\in N(\mathbf{p})} w_{\mathrm{BF}}(\mathbf{p},\mathbf{q})}, \tag{3}$$

whereas the weighting term w_{BF} in (2) considers the corresponding high-resolution 2-D image I. Thereby and in the following depth enhancement techniques by fusion, the low-resolution depth map D has to be upsampled to have the same spatial resolution as I. To this end, nearest-neighbour interpolation might be used [25,48] since alternative data interpolation approaches such as bilinear or bicubic interpolation might introduce non-valid depth measurements among the upsampled edges of the depth map [17]. As in (1), the resulting depth map $\mathrm{J}_{\mathrm{JBU}}$ is an enhanced version of D that has been upsampled to the 2-D guidance image resolution. Nevertheless, the violation of the heuristic assumption in which depth and colour are strongly correlated [2] may lead to erroneous copying of 2-D texture into actually smooth