ISPRS Journal of Photogrammetry and Remote Sensing 125 (2017) 16–32

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Joint inpainting of depth and reflectance with visibility estimation ☆

Marco Bevilacqua [a,b,c,d,*], Jean-François Aujol [a], Pierre Biasutti [a,b,c], Mathieu Brédif [c], Aurélie Bugeau [b]

[a] Université de Bordeaux, IMB, CNRS UMR 5251, 33400 Talence, France
[b] Université de Bordeaux, LaBRI, CNRS UMR 5800, 33400 Talence, France
[c] Université Paris-Est, LASTIG MATIS, IGN, ENSG, 94160 Saint-Mande, France
[d] Bordeaux INP, IMS, CNRS UMR 5218, 33400 Talence, France

ARTICLE INFO

Article history:
Received 22 July 2016
Received in revised form 3 January 2017
Accepted 5 January 2017

Keywords:
Inpainting
Total variation
Depth maps
Lidar
Reflectance
Point cloud
Visibility

ABSTRACT

This paper presents a novel strategy to generate, from 3-D lidar measures, dense depth and reflectance images coherent with given color images. It also estimates for each pixel of the input images a visibility attribute. 3-D lidar measures carry multiple information, e.g. relative distances to the sensor (from which we can compute depths) and reflectances. When projecting a lidar point cloud onto a reference image plane, we generally obtain sparse images, due to undersampling. Moreover, lidar and image sensor positions typically differ during acquisition; therefore points belonging to objects that are hidden from the image view point might appear in the lidar images. The proposed algorithm estimates the complete depth and reflectance images, while concurrently excluding those hidden points. It consists in solving a joint (depth and reflectance) variational image inpainting problem, with an extra variable to concurrently estimate handling the selection of visible points. As regularizers, two coupled total variation terms are included to match, two by two, the depth, reflectance, and color image gradients. We compare our algorithm with other image-guided depth upsampling methods, and show that, when dealing with real data, it produces better inpainted images, by solving the visibility issue.

1. Introduction

Image-based 3D reconstruction of static and dynamic scenes (Seitz et al., 2006; Herbort and Wöhler, 2011; Stoykova et al., 2007) is one of the main challenges in computer vision nowadays. In the recent years many efforts have been made to elaborate configurations and approaches, possibly requiring the employment of multiple sensors, with the final goal of generating plausible and detailed 3D models of scenes. To this end, typical optical cameras are often combined with non-visual sensors. The intermediate outputs of these hybrid systems, prior to the final scene rendering, are in general depth or depth + color images (RGB-D). Among the non-visual sensors, we can find Time-of-Flight (ToF) cameras (Kolb et al., 2010), which acquire low-resolution co-registered depth and color images at a cheap cost, and the famous Kinect (Zhang, 2012), capable to extract depth information by exploiting structural light. Another possibility is represented by lidar devices, which are used in a variety of applications and provide as output point clouds with measures of distance and reflectivity of the sensed surfaces.

This work lies in the context described and is particularly driven by the exploitation of data acquired by Mobile Mapping Systems (MMS), such as Paparoditis et al. (2012). MMS systems are vehicles equipped with high-resolution cameras and at least one lidar sensor: their contained dimensions allow them to be driven through regular streets and acquire data of urban scenes. The data acquired is a set of calibrated and geolocated images, together with coherent lidar point clouds. The interest towards them comes from the possibility of having available, at a relatively small processing cost, the combination of depth and color information, without having to perform explicit (error-prone) reconstructions. Having a good depth estimate at each pixel, for example, would enable the possibility to perform depth-image-based rendering algorithms, e.g. Zinger et al. (2010), Chen et al. (2005), Schmeing and Jiang (2011). Similarly, the availability of depth information allows the insertion of virtual elements into the image, such as pedestrians or vehicles generated by a traffic simulation (Brédif, 2013). While MMS data sets do not include directly depth images aligned with

☆ This study has been carried out with financial support from the French State, managed by the French National Research Agency (ANR) in the frame of the Investments for the future Programme IdEx Bordeaux (ANR-10-IDEX-03-02). J.-F. Aujol also acknowledges the support of the Institut Universitaire de France.
* Corresponding author at: Bordeaux INP, IMS, CNRS UMR 5218, 33400 Talence, France.

http://dx.doi.org/10.1016/j.isprsjprs.2017.01.005
0924-2716/

the available color images, it is easy, by exploiting the known geometry, to project the lidar point clouds onto each image. This operation produces initial depth images, which present three main issues (see Fig. 1, where three parts of an input depth image are shown, together with the corresponding image parts).

1. *Undersampling*: since lidar and image acquisitions are deeply different in terms of geometry and characteristics, the resulting depth images turn to be irregular. No points are present in the sky and on reflective surfaces. Moreover, the point density, which depends on the variable distances between the camera image plane and the positions of the lidar sensor, is generally significantly smaller than the pixel resolution. We can therefore talk about sparse input depth images (see for example Fig. 1a, showing the low density of lidar points from the ground).
2. *Visibility* (hidden parts appear): since points that are not visible from the image view point (hidden points) can be occasionally "seen" by the moving lidar sensor, erroneous values referring to such points can appear in the input depth image. This occurs even when a Z-buffer approach (Greene et al., 1993) is used, i.e. only the closest depth values for each pixel are kept (in case multiple values end up in the same pixel location). E.g., Fig. 1 b shows that depth values from the building behind appear as foreground points.
3. *Occlusions* (visible parts disappear): for the same reason as above, i.e. the different acquisition timing and geometry between image and lidar sensors, surfaces normally visible from the image view point do not get a corresponding depth. This can happen when the lidar sensor suffers occlusions at a given instant or because of the scene dynamics. E.g., in Fig. 1c, a moving bus that is not present at the moment of the image shot happens to appear in the depth image.

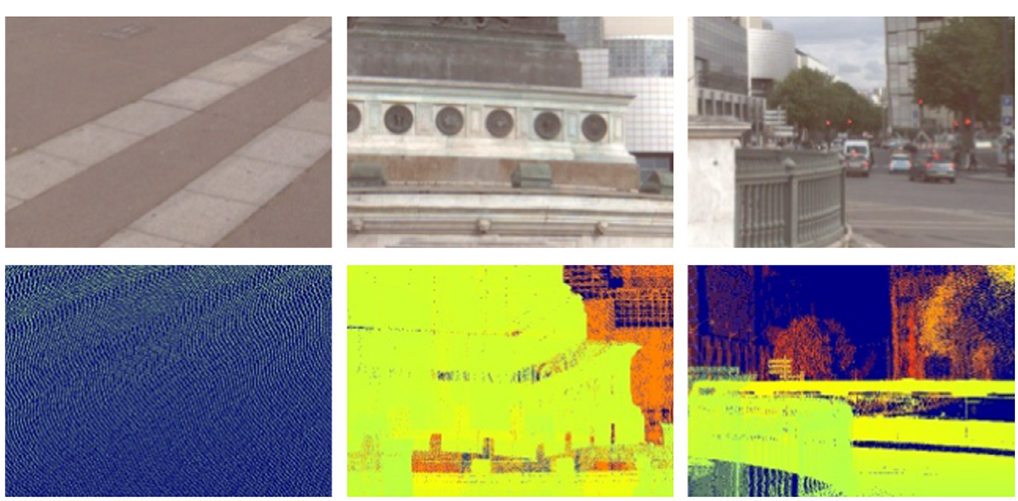
While there is a variety of methods in the literature that deal with the first issue, i.e. that aim at upscaling an irregular input depth image possibly with the guidance of a corresponding color image, little work has been performed to address the last two issues. In this paper, while inpainting the input depth image, we also intend to tackle the visibility problem. Moreover, we treat at the same time an additional input: a sparse reflectance image derived in the same way as the input depth image (i.e., by naively projecting the lidar point cloud, considering the reflectance information carried out by each point). We will show that the simultaneous use of a reflectance image, which is inpainted jointly with the depth, improves the quality of the produced depth image itself. To jointly inpaint depth and reflectance and concurrently evaluate the visibility of each point (i.e. establish if a single point is reliable or, since non-visible, must be discarded), we formulate an optimization problem with three variables to estimate: depth, reflectance and a visibility attribute per pixel. The inpainting process is also guided by the available color image, by means of a twofold coupled total variation (TV) regularizer.

The remainder of the paper is organized as follows. In Section 2, we present our approach and mention the related works, in particular on the image-guided depth inpainting problem. In Sections 3 and 4 we describe the model used and the primal-dual optimization algorithm that arises, respectively. Finally, in Section 5 we bring experimental evidence that proves the effectiveness of the proposed approach.

2. Problem addressed and related work

Fig. 2 depicts the scheme of the proposed approach. Given an MMS data set consisting of a lidar point cloud and a set of camera images, we choose among the latter a reference color image (w), and we obtain input depth (u_S) and reflectance (r_s) images by re-projecting the lidar points according to the image geometry. The two lidar-originated images are sparse images with irregular sampling and need to be inpainted. We propose to do that jointly and simultaneously estimate the visibility of the input points, within a variational optimization framework. The output of the algorithm are then three: the inpainted depth and reflectance (u and r, respectively), and a binary image expressing the visibility at each point (v).

In the literature there is a variety of methods that aim at upscaling or inpainting an original sparse depth image. Most of them are presented in the context of ToF cameras; thus, a high quality color image is acquired at the same time and can be exploited. We refer to this problem as image-guided depth inpanting. The typical assumption, when exploiting the available image, is that image edges are related to depth edges. Following this principle, many approaches have been proposed, such as methods using different versions of multilateral filtering (Chan et al., 2008; Yang et al., 2013; Garcia et al., 2010), methods based on Markov Random Fields (Diebel and Thrun, 2005), and methods using Non-Local Means (Park et al., 2011; Huhle et al., 2010). Another family relates



(a) Undersampling (b) Visibility (c) Occlusions

Fig. 1. Examples of parts from a resulting input depth image (bottom row), with the corresponding parts from the reference color image (top row), showing the three issues mentioned: undersampling, appearance of hidden points, and presence of occlusions.