Diffusion and inpainting of reflectance and height LiDAR orthoimages

Pierre Biasutti, Jean-François Aujol, Mathieu Brédif, Aurélie Bugeau

Diffusion and inpainting of reflectance and height LiDAR orthoimages

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

[a] *IMB, CNRS, IPB, University of Bordeaux, France*
[b] *LaBRI, CNRS, University of Bordeaux, France*
[c] *Univ. Paris-Est, LASTIG MATIS, IGN, ENSG, F-94160 Saint-Mandé, France*

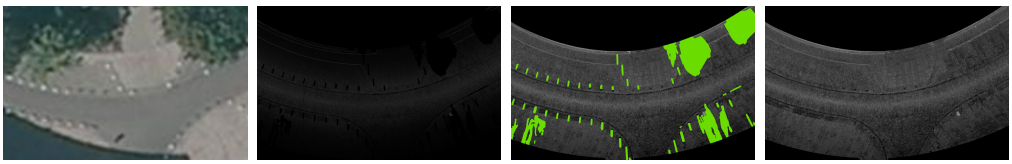
Abstract

This paper presents a fully automatic framework for the generation of so-called LiDAR orthoimages (*i.e.* 2D raster maps of the reflectance and height LiDAR samples) from ground-level LiDAR scans. Beyond the Digital Surface Model (DSM or heightmap) provided by the height orthoimage, the proposed method cost-effectively generates a reflectance channel that is easily interpretable by human operators without relying on any optical acquisition, calibration and registration. Moreover, it commonly achieves very high resolutions (1cm^2 per pixel), thanks to the typical sampling density of static or mobile LiDAR scans.

Compared to orthoimages generated from aerial datasets, the proposed LiDAR orthoimages are acquired from the ground level and thus do not suffer occlusions from hovering objects (trees, tunnels and bridges), enabling their use in a number of urban applications such as road network monitoring and management, as well as precise mapping of the public space e.g. for accessibility applications or management of underground networks.

Its generation and usability however faces two issues : (i) the inhomogeneous sampling density of LiDAR point clouds and (ii) the presence of masked areas (holes) behind occluders, which include, in a urban context, cars, tree trunks, poles or pedestrians (i) is addressed by first projecting the point cloud on a 2D-pixel grid so as to generate sparse and noisy reflectance and height images from which dense images estimated using a joint anisotropic diffusion of the height and reflectance channels. (ii) LiDAR shadow areas are detected by analysing the diffusion results so that they can be inpainted using an examplar-based method, guided by an alignment prior.

Results on real mobile and static acquisition data demonstrate the effective-



a. b. c. d.

Figure 1: a) Aerial orthoimage. b) Rasterized LiDAR pointcloud (reflectance attribute). c) Interpolated LiDAR refectance with estimated occlusion mask in green. d) Proposed LiDAR orthoimage with inpainted reflectance.

ness of the proposed pipeline in generating a very high resolution LiDAR orthoimage of reflectance and height while filling holes of various sizes in a visually satisfying way.

---

1. Introduction

Orthophotographies and Digital Surface Models (DSM), defined respectively as the color and ground height orthoimages (*i.e.* raster maps defined on a regular horizontal grid), are ubiquitous products in modern cartography. They are widely used in many application fields such as remote sensing, geographical information and earth observation, mapping and environmental studies. Such orthoimages are traditionnaly computed from an aerial perspective (satellites, planes and more recently unmanned aerial vehicles (UAVs)). Although aerial imagery approaches provide a very well known and common approach to the problem of orthoimage generation, they may be limited in terms of accuracy and resolutions and they certainly suffer from occlusions caused by the natural and urban environnement such as trees, tunnels, overhangs or tall buildings (Fig. 1.a).

These limitations prevent orthoimages generated by above-ground datasets to be used for a whole new set of applications that rely on a precise mapping of the ground and which cannot suffer from such large occlusions. These applications include, mostly in a urban context, accessibility assessment for soft mobilities (disabled, wheelchairs and strollers) and itinerary computations (39), precise mapping of road marks (24), road limits or curbs (29; 19; 51; 23), road inventory (33), road surface modelling and quality measurements (22), mobile mapping registrations on aerial images (43) or image based localization using ground landmarks (34). Moreover, recent legislations in European

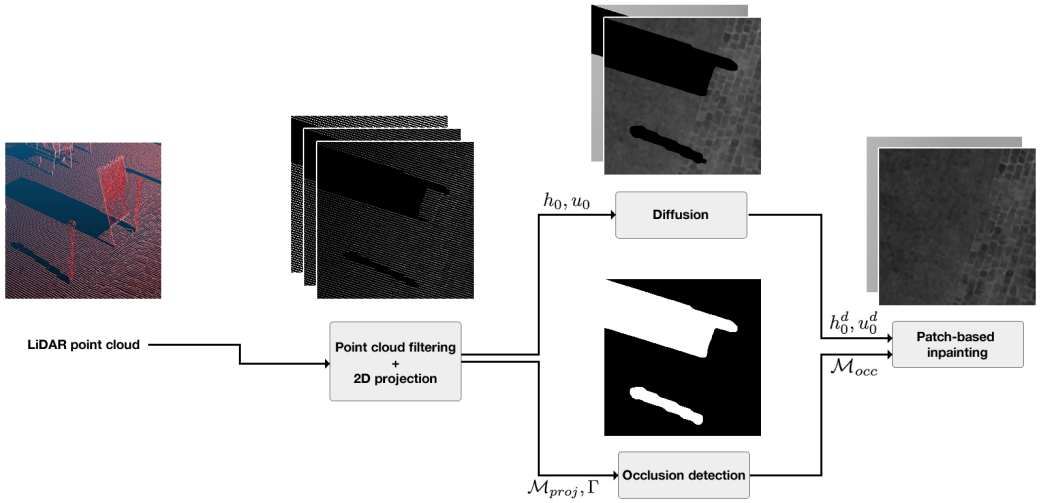


Figure 2: Full orthoimage production pipeline from MLS. Framed rectangles are processing steps, arrows are exchanged data. h_0 and u_0 are the projections of each point height and reflectance respectively onto an horizontal grid. $\mathcal{M}_{proj}$ is a binary mask of pixels where at least one point was projected.

countries call for a subdecimetric accuracy mapping of underground networks (water pipes, gaz pipes, internet wires and phone wires) as the lack of accurate data has lead to accidents and delays in many public works. Very high resolution orthoimaging with limited occlusions could help in meeting the requirements of these legislations as it would provide sub-centimetric accuracy mapping of the ground.

To maximize orthoimage resolution and to minimize occlusions, we propose to leverage ground-level LiDAR scans acquired by Mobile Laser Scanning vehicles (MLS) or from fixed stations. The proximity of the acquisition ensures a high resolution as well as a diminution of occluded areas. As in (44), we propose to derive a gray image from the reflectance attribute of the LiDAR samples (which measures backscattered energy) instead of relying on optical imagery (which would introduce difficulties in dynamic environments and require precise co-registration) to produce sub-centimetric orthoimages.

1.1. Problem statement and related works

The projection of a ground-level point cloud at centrimetric resolutions creates a sparse image due to its inhomogeneous sampling density (Fig. 1.b).

The problem of filling in the holes created by the lack of information in the raster image created by the projection of a point cloud has been scarsely investigated (44). This problem is strongly related to DSM generation from LiDAR data, especially airborne LIDAR data, which has been widely studied over the past decades (15). DSMs are mostly represented either by Triangular Irregular Network (50; 20; 14) or by raster images (27; 45; 41; 13). In both cases, the main challenges of DSM generation still remains ground point filtering and interpolation. Ground point filtering aims at removing points that belong to the ground (*e.g.* earth surface in case of airborne DSMs) from points that belong to elevated structures (such as buildings, trees, cars, fences or poles). For airborne LiDAR data, it is done by defining slope operators in order to follow the ground surface (49; 42; 25). However, these methods are developped in order to extract the ground on large scale, for terrain with high relief variation. In the case of urban scenarios, these methods fail to distinguish correctly small objects (bikes, pedestrians) from the ground. DSM interpolation approaches depends on the final product representation. For TIN, the interpolation is either done by Delaunay Triangulation (48) or by plane fitting, as proposed in (7). However, these approaches are not relevant to our problem as we aim at generating orthoimages as well as raster DSMs. The interpolation of raster images DSMs has already been done by coarse to fine interpolation approach (27), by using moving least squares in (41) and more recently using Poisson interpolation (31) to deal with the high variations of density in the raster images derived from MLS (44). The generation of DSMs does not require to preserve textures as the surface model is textureless. However in our context, we aim at generating orthoimages from the reflectance as well as DSMs. Therefore, the preservation of texture is a key point of our problem, which requires to use other approaches for the raster image interpolation.

1.2. Contributions

We propose a framework that combines diffusion inpainting and examplar-based inpainting for the joint production of high resolution and photo-realistic reflectance and height orthoimages. Our main contribution is to provide a fully automatic framework with an efficient set of default parameters. First, we introduce a novel anisotropic diffusion technique that uses both texture and height information. Second, we propose an examplar-based inpainting method that is modified in order to take advantage of the specific attributes present in LiDAR datasets.

1.3. Outline of the paper

In this paper, we propose in Section 2 a novel approach for the reconstruction of ground orthoimages that produces very satisfying results in terms of visual quality and coherence. We first explain in Section 3 how the point cloud can be projected on a 2D-pixel grid after filtering ground points. We present a brief state-of-the-art on diffusion algorithms and we introduce a diffusion model for the computation of dense images from the sparse projections in Section 4. In Section 5, we present different examplar-based inpainting techniques before introducing an inpainting method that takes height information into account as well as assumptions about the alignment between structures. Finally, in Section 6, we validate our framework by presenting various examples of reconstruction made on real acquisition data.

2. Framework description

Orthoimage generation from LiDAR scans aquired at ground level as been scarcely studied in the past. Nevertheless, the relation between LiDAR reflectance and optical acquisition has already been used for different applications such as depth map generation from point cloud (6), which shows the correlation between both modalities. (44) first extract the ground points by considering the lowest points projected in each pixel. Next, they perform Poisson interpolation to connect sparse pixels. In this paper, we introduce an efficient and fully automatic pipeline to reconstruct an aerial image from a LiDAR point cloud. The proposed framework is summed up in figure 2.

From the point cloud, we need to extract the ground points (eg. points that do not belong to a mobile object or to an object that is lying on the ground). This is done by computing an envelop Γ (see section 3). The reflectance and height values of these ground points are then projected in two 2D-images respectively: u_0 and h_0. This projection is done by removing the z (height) coordinate and rounding the coordinates to the chosen resolution. We also build a mask $\mathcal{M}_{proj}$ of the pixels where at least one point was projected.

At this point, the projections u_0 and h_0 are sparse as they do not cover all the pixels of the images. Figure 3 presents an example of the different kinds of missing pixels that results from the projection. Some parts of the projection correspond to the inside of a building (figure 3.c in orange), under sampling holes appear in between lines of acquisition (figure 3.c in blue) and an occlusion is caused by a pole blocking the laser beams (figure 3.c

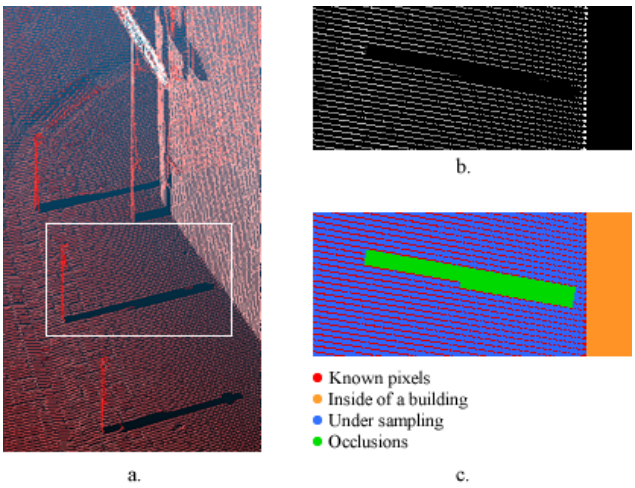


Figure 3: Highlighting of the different types of holes in the sparse projection. (a) is the original point cloud colored with the reflectance, (b) is the sparse projection of the white rectangle in (a) after extracting ground points, (c) is the sparse projection labelled with the different kinds of holes.

in green). In order to reconstruct the missing information of the orthoimage, we first perform diffusion on both u_0 and h_0 by coupling reflectance and height in an anisotropic diffusion algorithm in order to remove holes due to undersampling. The resulting images are respectively called u_0^d and h_0^d. After this step, there are still some large holes remaining. Their locations $\mathcal{M}_occ$ are retrieved through mathematicl morphology. Finally, we can reconstruct occlusion holes using an examplar-based inpainting method that uses both reflectance and height information, as well as an assumption about the alignment between structures to inpaint.

3. Projection of LiDAR point cloud

The projection of a point cloud onto a 2D pixel grid is a typical discretization problem. It mostly requires to define a mapping between the point cloud metric frame and the 2D-pixel grid. However, in the case of Digital Terrain Model, it is also needed to filter out off-ground points (trees, urban structures, cars). We introduce a novel approach for ground point filtering in section 3.1 and explain how the projections are done in section 3.2. More details about the parametrization of the projection can be found in 3.3.

3.1. Filtering ground points

The definition of ground-points in a point cloud can be tedious as we have to filter groups of point that represent relatively planar structures and which do not belong to any other objects than the ground itself. Ground filtering is a typical DTM generation problem (30). Traditional aerial DTMs generally model the scene at large scale. In order to correctly include details of urban scenes (pavements, steps or any lightly elevated structure that belongs to the ground), it is necessary to model the ground at a finest scale. In urban scenario, plane fitting is often use as primary ground segmentation. Although it allows a fast and simple estimation of ground points, considering horizontal planes relatively to the acquisition system can be ambiguous. Indeed, modern MLSs tend to be accurate enough to acquire ceilings through windows, creating false positives. Vertical planes are also relevant (pavements, stairs), but not in every cases (trucks, billboards). This problem has been investigated by considering it as a classification problem (36) or by performing advanced structural analysis (27; 9). However, these solutions have shown their limitations when the scene presents high diversity of objects. In particular, they lack of precision when aiming at estimating the boundaries of the ground in urban scenes because other objects (cars, ceilings) are often considered as the ground as they share common structural properties.

We propose a novel approach for ground point filtering based on the way the acquisition is done. We aim to filter out hovering object or any point that is over another one. As the points are acquired with a certain margin of error, direct comparison is not suitable as the likelyhood of two points having the exact same (x, y) coordinates is negligible. We first create an empty envelop of the size of the projection where each pixel has an infinite value. This envelop will help defining the boundaries of the 2D region that represents the ground while ensuring that all the points that fall into the envelop really are ground points. We then consider segments made by each point and its relative emission point. Each segment is discretized in the envelop using the Bresenham line algorithm (10). As the beam is perfectly straight, we can estimate the height of the segment at any position of the segment. Each pixel is then filled with the lowest height value of segments that cross it. Note that in our case, only points below the sensor are considered. This reduce the amount of data to process while ensuring to discard only off-ground points. However, this is only suitable for MLS in urban scenarios. Figure 4 shows a slice of the maximal envelop Γ computed on a set of beams that overlaps. We can see that for every overlapping beams, only the portion of the lowest

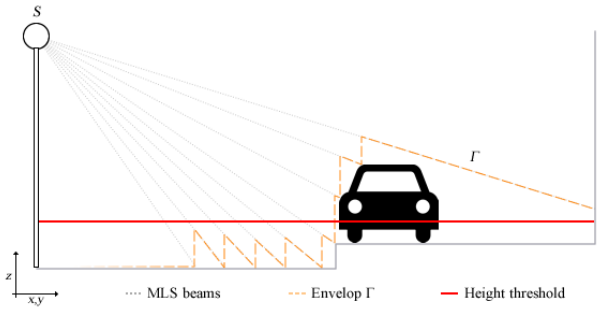


Figure 4: Slice of an envelop Γ obtained by evaluating several aligned beams coming from the sensor S until they hit an object of the scene. The red line is the final threshold applied to the envelop to exclude too high points.

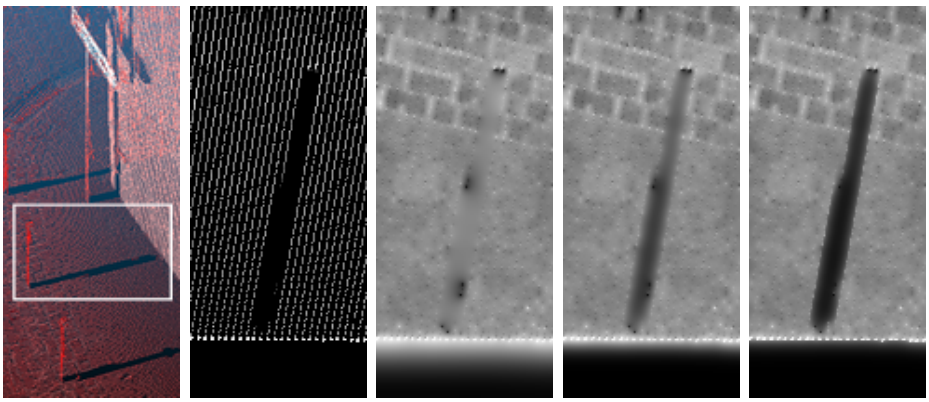
one is kept in the envelop. Finally, we filter the point cloud by taking only points that are under the envelop and the threshold, with an epsilon margin.

3.2. Sparse projections

Using the filtered point cloud, we want to produce two sparse images corresponding to the reflectances and the heights in the sensor frame: u_0 and h_0 defined on the mask $\mathcal{M}_{proj}$. The values for each pixel in u_0 is the mean of the reflectances of every points that is being projected in it. The values in h_0 are the same using the height in the sensor frame. Finally, an $\mathcal{M}_{proj}$ image is produced where pixels are valued 1 where at least one point was projected and 0 elsewhere. Note that at high resolution (1px per square centimeter), the use of the mean is relevant on our data as only few points (less than 5) project in each pixel. However, if the amount of points that projects in each pixel increases a lot (when working at lower resolution for example), one can consider using the median instead of the mean to remove outliers. Note that the computational cost of the median is higher than the cost of the mean. Thus, its use will significantly increase the running time of the projection step.

3.3. Parameters

The choice of the mapping between real coordinates and pixels mostly depends on the density of the point cloud. In our case, with an acquisition done using a RIEGL LMS-Q120i which produces 300 000 points per second, the maximal acceptable resolution was 1px = 1cm^2. The height threshold



a b c d e

Figure 5: Comparison of different diffusion techniques for filling stripe holes. (a) is the point cloud, (b) its projection (rotated for clarity purpose), (c) is the Gaussian diffusion result, (d) is the Perona-Malik algorithm result and (e) is the result of our proposed modification. We can see that our modification provides a better conservation of big holes while filling perfectly the stripe holes.

is arbitrary but in the case of a urban scenario, it should be kept under the height of the aquisition vehicle. More details about the parameters are provided in 6.1.

3.4. Dependency to the sensor

It is important to point out that the type of missing data are directly related to the chosen resolution as well as the type of sensor. The holes due to the acquisition sampling are less likely to appear when chosing a lower resolution. Moreover, the missing values in between acquisition lines are specific to the sensor mentionned above. They are quite homogeneous and create a regular pattern. With a panoramic sensor such as the one used in (1), the missing pixels will appear in a random pattern, but will create a more dense image for the same resolution, which makes our pipeline still suitable for this type of data.

4. Diffusion of sparse images

The two images obtained in the previous section are sparse in the sense that they do not cover every pixels of the DTM. Therefore, we need to interpolate the images in order to get a dense representation of them. The goal is

to fill in gaps between relatively close pixels that are due to the acquisition undersampling. In this section, we first explain what are the requirements that the filling method needs to meet. Then we introduce a modification to existing methods in order to enhance the results. Finally we show a comparison of different methods to validate our proposed modification.

4.1. Choice of the approach and requirements

A typical approach for filling small holes by interpolation is to use diffusion algorithms. Several diffusion techniques exist such as the total variation (12), the generalized total variation (8), structure tensor diffusion (47; 5) or partial differential equation diffusion (2) and extended to multi-modal data (52).

Here, we focus on iterative solving methods which are more flexible. A basic diffusion algorithm is the so called Gaussian diffusion which is an isotropic technique that consists in updating the image with its own Laplacian (26). However in the case of a urban scenario, an anisotropic diffusion is more relevant as very high gradients appear at the edge of different structures (roads, pavements, stairs) and need to be preserved.

The Perona-Malik algorithm (32) is a well known algorithm for anisotropic diffusion. It is partially inspired from the Gaussian diffusion and is defined as follows:

$$\begin{cases} \frac{\partial u}{\partial t} - \mathrm{div}(c(|\nabla u|)\nabla u) = 0 \text{ in } \Omega \times (0,t) \\ \frac{\partial u}{\partial N} = 0 \text{ in } \partial\Omega \times (0,T) \\ u(0,x) = u_0(x) \text{ in } \Omega \end{cases} \tag{1}$$

where $u_0 \in \Omega$ is the input image, *div* is the divergence operator, ∇ is the gradient operator, N is the normal vector to the boundary of Ω and c is an increasing function. A common choice for c is the weighting function $c(|\nabla u|) = \frac{1}{\sqrt{1+(|\nabla u|/\alpha)^2}}$, α being a weighting factor that quantifies how much the gradient information needs to be considered. This technique ensures the preservation of edges while ensuring smooth transitions between sampled scan lines. Nevertheless, this technique only takes into account the gradients of a single channel. In our context, the diffusion needs to be blocked in case of a high gradient in the reflectance image as well as in the case of a high gradient in the height image that could correspond to the junction between the road and a pavement, or steps of stairs. Therefore, we need to modify equation (1) in order to take both channels into account.

4.2. Our proposed algorithm

We propose here a modification to the Perona-Malik equation (1) by coupling heights and reflectances as follows, using previously introduced notations:

$$\begin{cases} \frac{\partial u}{\partial t} - \text{div}(f(|\nabla u|, |\nabla h|)\nabla u) = 0 \text{ in } \Omega \times (0,t) \\ \frac{\partial h}{\partial t} - \text{div}(f(|\nabla u|, |\nabla h|)\nabla h) = 0 \text{ in } \Omega \times (0,t) \\ \frac{\partial u}{\partial N} = 0 \text{ in } \partial\Omega \times (0,T) \\ \frac{\partial h}{\partial N} = 0 \text{ in } \partial\Omega \times (0,T) \\ u(0,x) = u_0(x) \text{ in } \Omega \\ h(0,x) = h_0(x) \text{ in } \Omega \end{cases} \tag{2}$$

where we recall that u_0 is the reflectance image and h_0 is the height image. We introduce the new weighting function f that emerges from the one used in equation (1) as follows:

$$f(|\nabla u|, |\nabla h|) = \frac{1}{\sqrt{1 + \frac{|\nabla u|^2}{\alpha^2} + \frac{|\nabla h|^2}{\beta^2}}} \tag{3}$$

having α, β as weighting constants quantifying how gradients of reflectance and height need to be considered. The choice of coupling both reflectance and height information into the same model is motivated by the fact that reflectance and height gradients are not always at the same locations and therefore, are complementary. Note that coupling various modalities in a model has already been proposed in (3) for coupling multi-spectral images, however in that case authors present a model specifically designed for merging multiple images representing the same object at different wavelengths. Using our method, we can now take into account gradients coming from both u_0 and h_0.

4.3. Comparison with other diffusion techniques

In this section, we propose an evaluation of the performances of our model against gaussian diffusion and closest neighbors diffusion. Projecting a point cloud acquired at very low speed provides a dense image locally. Therefore, we can define a ground truth using this region of the projection. We define a set of 20 masks of same dimension as the ground truth and we randomly set 80% of the pixels to 1. For each method and each mask, we recover pixels of the ground truth where the mask is valued 1, using the rest of the

a b c d e

Figure 6: Result of the different diffusion models on degraded ground truth. (a) original image, (b) original image with 80% of pixels removed, (c) Closest Neighbors result, (d) Gaussian diffusion result and (e) Our result. We can see that the Gaussian diffusion and our model better recover the aspect of the image. Our method succeed in a finer edge recovery.

Table 1: Evaluation of different diffusion algorithms

Metric	Closest Neighbors	Gaussian	Proposed Model
MSSIM	0.8056	0.8550	**0.8591**
MPSNR (dB)	33.21	34.59	**35.08**

image. Note that the percentage of missing pixel (here, 80%) is defined as the average missing pixels ratio of our dataset. Finally, we compute the average of classical similarity metrics (MSSIM and MPSNR, which are respectively the mean of the SSIMs and the mean of the PSNRs) for each methods on the reconstructed images compared to the ground truth. The results are summed up Table 1 in which we can see that our method outperforms the two other diffusion methods. Figure 6 presents one set of results. We can see that the Gaussian model as well as our model better succeed in recovering the aspect of the original image. Our method outperforms the Gaussian diffusion by recovering sharper edges.

4.4. Parameters

In practice, the proposed diffusion technique was implemented by solving the PDE system with a first order explicit Euler scheme with respect to the time variable. The number of iterations has to be chosen in order to fill in stripe holes. It depends on the chosen resolution as very sparse images will require more iterations to fully fill the image. Moreover, a good speed-up can be obtained by using the result of the closest neighbors diffusion of both u_0 and h_0 as the initialization for the proposed model as it drastically lowers the number of required iterations. The weighting term for the reflectances α should be higher than the one for height β in order to completely block the diffusion in case of large height variation while connecting close pixels. Practical details will be given in section 6.1. Note that only unknown pixels

regarding $\mathcal{M}_{proj}$ should be updated to prevent an oversmoothing of the final images.

5. Inpainting of occlusions

After the projection, some holes are not only caused by some undersampling but also by the beam being blocked by an object (cars, poles, lights, pedestrians or bikes) before reaching the ground. This leads to a ground projection with a lot of information at the edge turned toward the sensor, but nothing when going further. As occlusion holes are wider than stripe holes, the diffusion algorithm proposed above is not suitable in order to reach a visually satisfying result. In this section, we first see how occlusions holes are detected in the image. We then present the problem of texture synthesis in our case and we give a first solution. Finally, we introduce an improvement to this solution based on assumptions made on the urban scenario.

5.1. Occlusion hole detection

The occlusion detection consists in defining which holes are caused by the sampling rate and which holes are caused by a blocking of the laser beams. This can be done by applying mathematical morphology on the projection mask $\mathcal{M}_{proj}$ before diffusion where each known pixel is valued 1 and all other pixels are valued 0. At this point, everything with the 0 value is considered as occlusion holes.

Having $\mathcal{M}_{proj}$, a simple morphological operation known as closing (40) is enough to detect occlusions and build the occlusion mask $\mathcal{M}_{occ}$. The closing consists in applying a dilation of a certain radius to the mask and then to apply an erosion of the same radius. This leads to a closing of small 0-labelled areas surrounded by 1s. Choosing wisely the radius of the closing ensures that undersampling holes are eliminated while preserving the shape and the position of the occlusion holes.

Unfortunately, the resulting mask does not consider the boundaries of the scene, and tends to extend further. We recall that when projecting the point cloud (Section 3), a Γ envelop is computed in order to define the boundaries of the scene. Thus, we consider the intersection of the computed mask and the Γ envelop to prevent the mask from expending outside of the ground region, typically inside of buildings or in regions too far from the sensor (Figure 3).

5.2. Examplar-based inpainting

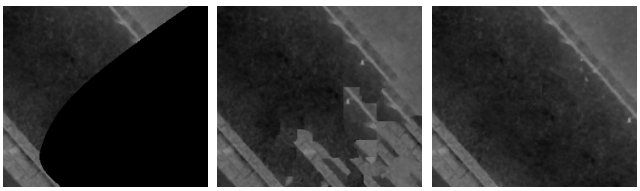
Among the variety of different inpainting algorithms, examplar-based algorithms are known for being more effective and more reliable in filling large areas (with large internal radius). Examplar-based inpainting consists in trying to find the best candidate in the known region of the image for the patch centered on a pixel lying on the border of the hole. Once found, the candidate is used to fill the unknown part of the image by copying the color in its central pixel (18) or the full patch (16). The operation is repeated until the hole is fully closed. More recent approaches, such as (17; 46) reconstruct the texture using both color information and depth information. However these algorithms require different acquisitions of the same view, which is not applicable in our case as we aim at performing the reconstruction on a single acquisition pass.

The urban scenario presents a huge variety of structures (roads, pavements, stairs, gutters) as well as many different textures (roads, cobbles, floor tiles). Thus, we decided to base our work on the Criminisi et al. (16) algorithm that was designed for the good preservation of the structures in the reconstruction. More complex approaches exist that rely on the work presented in (16) such as (11) and (28) however it would have been less intuitive to adapt them to our context. In (16), authors put forward the idea that the order in which areas are reconstructed have a high impact in the final result. They introduce a priority term that takes into account the strength and the direction of the image's gradient at the border of the unfilled area. A patch that contains a strong gradient in the direction orthogonal to the border of the region to reconstruct is evaluated before more uniform patches.

5.3. Modification to the original algorithm

Coupling reflectances and heights. The algorithm presented in (16) offers a very good technique for region filling. However, it can fail when the area to fill is very large. Therefore, we introduce a modification to the algorithm by taking the height information into account as a guide for the reconstruction. The idea is to use the height information to restrain the selection of best candidate patches to the areas of similar height by computing the Sum of Squared Differences of the candidate patch in both the reflectance and the height images. The SSD (Sum of Squared Differences) is defined as follows:

$$\mathrm{SSD}(\mathrm{P}_1, \mathrm{P}_2) = \sum_{\mathrm{i,j} \in \Omega} (\mathrm{P}_1(\mathrm{i,j}) - \mathrm{P}_2(\mathrm{i,j}))^2 \tag{4}$$



a. b. c.

Figure 7: Comparison of (16) and our proposed modification on the junction between the road and a pavement. (a) is the original unfilled image where the dark region is being reconstructed using examplar-based inpainting, (b) the result from (16), (c) our proposed optimization. The result is clearly better in (c) as the reconstruction conserves the structures of the image without creating new artifacts such as the one appearing on the left of (b).

having P_1, P_2 the two 2D-patches that are compared and Ω the domain of definition of the image. In our modification, and for each candidate, a score is attributed by combining both channels as follows:

$$S_p(\mathrm{P_t}, \mathrm{P_c}) = \mathrm{SSD}(\mathrm{P_t^R}, \mathrm{P_c^R}) + \eta \times \mathrm{SSD}(\mathrm{P_t^H}, \mathrm{P_c^H}) \tag{5}$$

where $\mathrm{P_t}$ is the target patch to be filled and $\mathrm{P_c}$ is a candidate patch. $\mathrm{P_c}$ can be any patch in the image that has no pixel that belongs to an occlusion hole. However, for speed-up purpose, we can limit the selection of $\mathrm{P_c}$ to be in a certain radius around $\mathrm{P_t}$. η is a regularization parameter and the superscripts R, H denote that the patch is taken in the reflectance image or the height map respectively. The regularization parameter only appears for the height map SSD as the height information is less important in order to reach a visually satisfying result.

The impact of the use of the height map in the synthesis is very noticeable in figure 7. The structure of the road is well preserved using the proposed modification compared to the original algorithm in which artifacts appear after some iterations. These artifacts mislead the reconstruction and the result is visually incoherent.

Taking advantage of urban environment. Although the current modification of the algorithm provides a very good solution for filling occlusion holes, the reconstruction can fail sometimes when the hole is very large. This happens for holes that are caused by cars or trucks where the area to reconstruct is significantly larger than regular holes (10^6 pixels at a $1px = 1cm^2$ resolution for a standard car and the portion of pavement behind it) and it can become

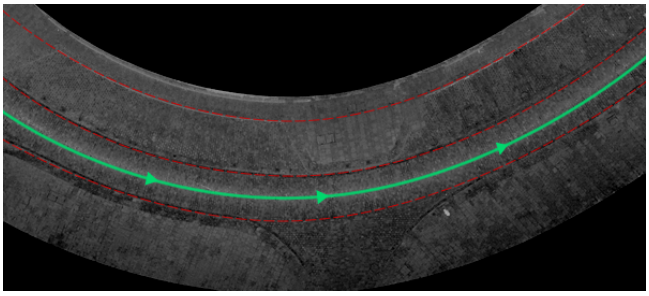


Figure 8: Illustration of the assumption that the urban environment evolves in a similar way than the path of the sensor. The straight green line shows the path of the sensor. Each dashed line represents areas of same distance to the sensor.

a common issue. Indeed, at the center of the holes the nearest known information is too far away and the error accumulated along the iterations is likely to fail the reconstruction. To improve the results in the concerned areas, we advocate that the structure of a urban environment is very likely to evolve in a similar way to the vehicle path as illustrated in figure 8. Therefore, we can constrain the selection of candidates to patches that are at a similar distance to the sensor than the current patch. The range attribute of the LiDAR image provides this information for each point.

We define the new score equation as follows, using previously introduced notations:

$$S_f(\mathrm{P_t}, \mathrm{P_c}) = \left[1 + \left(\frac{|\mathrm{d}(\mathrm{P_t}) - \mathrm{d}(\mathrm{P_c})|}{\gamma}\right)^2\right] \times \mathrm{S_p}(\mathrm{P_t}, \mathrm{P_c}) \tag{6}$$

having $d(P)$ the distance between the sensor and the center of the patch P and γ a regularization parameter that constrains the selection of patch to a range interval around the current range. The range can be accessed everywhere in the image by precomputing a signed distance map of the area to the path of the vehicle (eg. where the range is the lowest).

Large patches and artifacts. When the reconstruction is done at a very high resolution, large patches (10^3px) are likely to be required in order to correctly represent the structural elements of the image. This might lead to abrupt junctions between reconstructed patches. Therefore, we propose to enhance the copy of the patch by performing the seam carving using graphcuts presented in (37). The goal is to compute the optimal cut between $\mathrm{P_t}$ and $\mathrm{P_c}$ where they overlap to obtain a seamless result.

5.4. Parameters

η should be kept under 1 to ensure the visual coherence of the reconstruction. Parameter γ depends on the size of the occlusion. When $\gamma = 1$, the regularization is very strong and the selection of the candidate patch is constrained on a narrow band of same distance to the sensor point. When the value of the parameter is highly increased ($\gamma > 10^4$), no regularization operates and the algorithm behaves as if the range was not taken into account. Therefore, one can alternate between these two values for γ depending on the internal radius of the occlusion (see next section).

6. Results

We conclude this paper by presenting different results obtained using the proposed framework. We first present a general set of parameters for an automatic reconstruction of a set of orthoimages. We then demonstrate the efficiency of the solution by showing various results and comparison to existing methods. After that, we validate the quality of the framework using numerical criterions. Finally, some details about the computation time are drawn.

6.1. Parameters

In the same way as other pipelines, this one comes with a set of parameters that was used for producing every images displayed in this paper.

Projection. The objective of this study was to provide very high quality orthoimages. Therefore, all reconstructions were done at the maximal possible resolution of our sensor: 1px = 1cm^2. A threshold of 60cm from the road level was used to filter out points after the computation of the envelop.

Diffusion. For the diffusion step, we found the best balance of results by setting $\alpha = 5, \beta = 0.7$ with 3 iterations and by first interpolating u_0 and h_0 using the nearest neighbor algorithm.

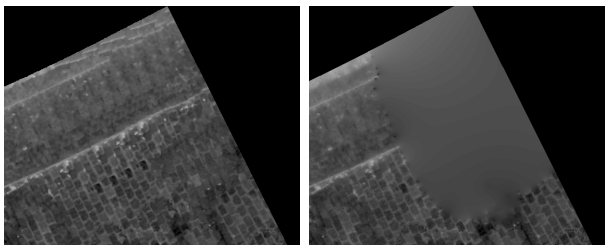
Mask extraction. In this step, a closing radius of 6px was enough to fill stripe holes while leaving occlusions intact.

Figure 9: Comparison between aerial orthophotography with a standard resolution (50cm^2 per pixel) (top) and MLS orthoimage using our model at 1cm^2 per pixel (bottom). Traditional orthophotography provides limited resolution and suffers from occlusions brought by the coverage of trees and other structures whereas our model provides unobstructed, hight resolution orthoimages. The aerial image comes from Geoportail.

Inpainting. At 1px = 1cm^2, the chosen patch size was 43x43px to fit the smallest structuring element (cobbles). In all our experiment, $\eta = 0.2$ ended up being a very good choice. Finally, we set the value of γ to 0.3 or 10^6, the choice being made by automatically checking whether the internal radius of the evaluated occlusion was higher than 50cm or not.

6.2. Qualitative analysis

A quick glance at the difference between traditional aerial orthophotography and MLS orthoimage using our framework is given in figure 9. The resolution provided by a typical aerial camera is about 50cm^2 per pixel, where our reconstruction is done at 1cm^2 per pixel. Fine textures and very precise details are noticeable in the reconstruction whereas only main structures can be seen in the aerial orthophotography. Moreover, the aerial orthophotography presents various occlusions such as trees that do not appear in our result.



a. b.

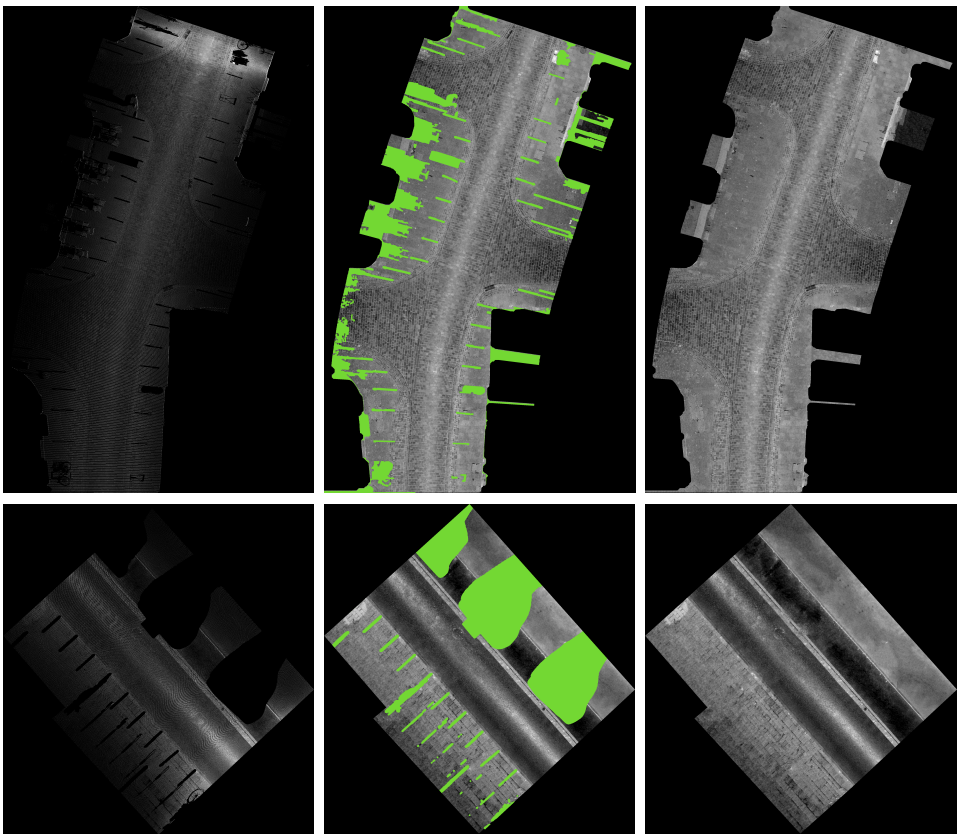
Figure 10: Comparison between our proposed framework (a) and the one introduced in (44) (b). Texture is better preserved using our framework.

In figure 10, we show a visual comparison between the proposed framework and the method introduced in (44) which is the state of the art for the production of orthoimages using MMS LiDAR point cloud. We can see that both algorithms perform about the same for stripe holes, but our solution gives more satisfying results for large occlusions. The texture is better reconstructed using our method. This will be later discussed in Section 6.3.

More reconstruction results are displayed in Figure 11. Each step of the pipeline is illustrated. We can see on Figure 11 top that the framework performs a very good reconstruction on fine details such as cobbles. In Figure 11 bottom, 25% ($\sim 5.10^5$px) of the area is occluded, mostly due to the presence of cars and poles. However, our framework succeeds plausible reconstruction of the scene, leading to a result that is much more understandable than initially. Finally, figure 12 shows an extreme scenario where the use of the range is relevant as the structure of the scene follows the same path as the road. The environment is fully reconstructed (16%, $\sim 10^6$px) while preserving the structure of the road.

In Figure 13, the framework is applied on data provided by the Semantic3D dataset (21). This dataset is aquired using a static LiDAR sensor. There, we can see that the area under the sensor as well as occlusion on the ground are successfully recovered while preserving the fine cobble texture.

The purpose of this pipeline is to generate both reflectance and height orthoimages. In figure 14, we show how the two outputs can be combined in order to obtain a 3D model of the road. Figure 14.a and 14.b are the reflectance image and the height image of the area that is being modelled in figure 14.c. We can see that the 3D model respects the topography of the



a. b. c.

Figure 11: Various results on different urban scenes. (a) shows the original point clouds projected on an horizontal grid (sparse). (b) are the results after stripe holes were filled. Areas that present large occlusions are highlighted in green. (c) the final results of our method. In both result, the orthoimage is successfully reconstructed while improving the understandability of the scene.

scene with the junction of the road and a pavement.

6.3. Quantitative analysis

Apart from the visual results, we also provide a numerical comparison between the proposed framework and the one of (44). Measuring similarities between two images is a tough task as the plethora of different metrics are all designed for a single aspect of the image (color variation, gradient similarity and correlation). In the case of texture synthesis, the similarity cannot be

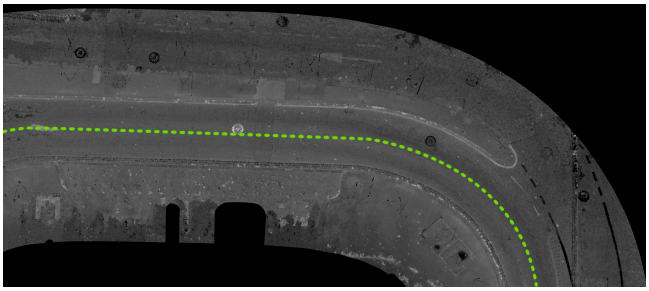


Figure 12: Exemple of scene that follows the vehicule path. In this case, the use of the range information is very relevant. The green dashed line denotes the vehicle path.

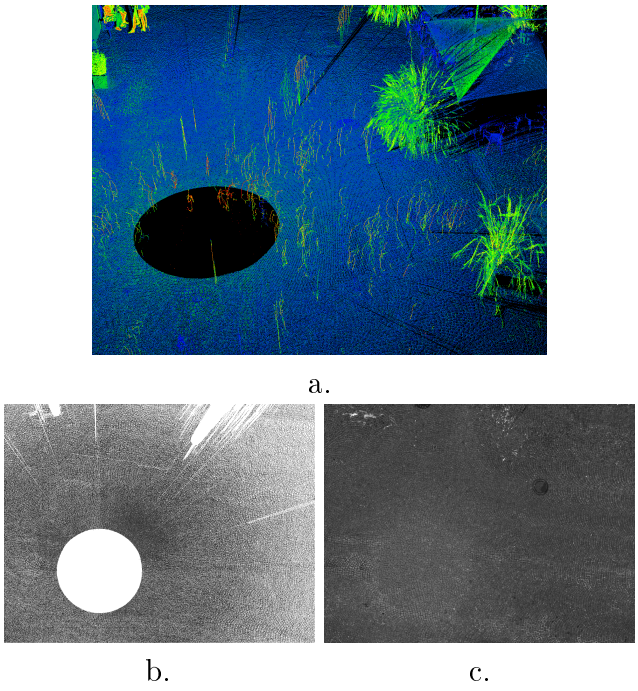


Figure 13: Example of reconstruction on the Semantic 3D dataset. (a) is the area from which the orthoimage is aquired, (b) is the projection of the ground points on an horizontal grid, (c) is the final result. The final results provides a plausible estimation of the area under the acquisition sensor.

directly compared as the goal is not to obtain exactly the same result, but to obtain visual coherence in the reconstruction. Thus, we advocate that the

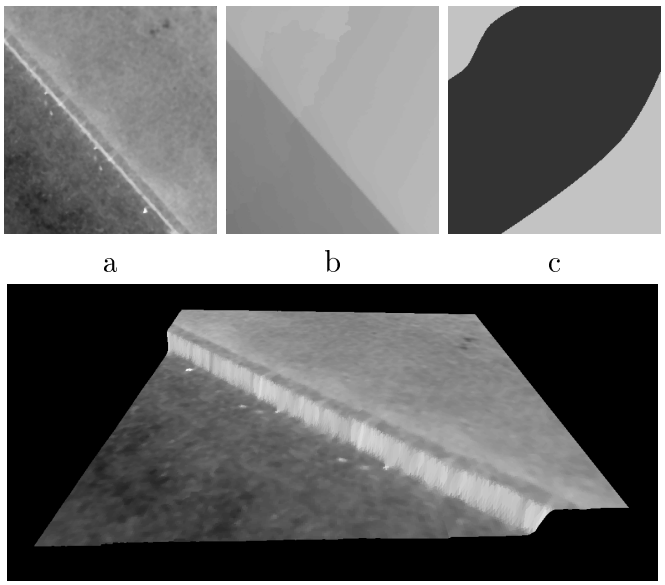


Figure 14: 3D model of the ground of a part of an orthoimage. (a) is the reflectance image used as texture for the 3D model, (b) is the height image used as the height coordinate of the 3D model, (c) is the mask where the darkest region was reconstructed using examplar-based inpainting, (d) is the 3D model obtained using both reflectance and height orthoimages.

Table 2: Numerical comparison between reconstructions

Image	Artificial occlusion		Real occlusion	
	STD	Hist. dist.	STD	Hist. dist
Ground truth	**4.51**	-	**4.79**	-
Proposed framework	**4.56**	**0.14**	**4.29**	**0.19**
(44)	1.87	0.78	2.05	0.80

measure of the standard deviation and the distance between histograms, also known as Wasserstein metric in (35), provide simple and efficient metrics for evaluating the quality of our results.

Table 2 sums up the comparison of the inpainting step on two examples: an image where the hole has been manually removed and an image where the ground truth is available as the vehicle did a second pass in which the occlusion disappear. For each example, we compute the standard deviation of the region reconstructed by examplar-based inpainting. We also compute the distance between the normalized histrograms of the ground truth and

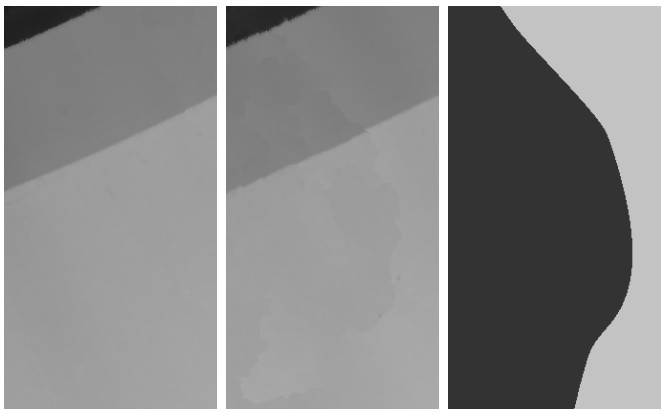


Figure 15: Comparison between height images with and without occlusion on the junction between a road and a pavement. (a) is the original height map, (b) is the reconstruction of an occlusion in the same area. The occlusion corresponds to the darkest region of (c). The mean square error of the reconstruction (b) compared to (a) on the occlusion region is 2mm.

each output. For both examples, our method provides a standard deviation that is very close to the ground truth resulting in visually similar textures.

As the proposed framework also reconstructs the height map of the aquired area, we provide a numerical analysis of this aspect. The choice of the metric in that case is quite easier as the height map is more homogeneous than the reflectance image, especially in a urban scenario as can be seen in figure 15. Therefore, the Normalized Mean Square Error is enough to estimate how good the reconstruction is. We found out that in general the mean square error was below 1cm. This validates the proposed framework for the reconstruction of height map.

6.4. Computational speed

The performances of the framework in terms of computational speed are mostly affected by the amount of occlusions and the resolution at which the reconstruction is being made. As the framework is composed of several steps, we present the computation time of each step as well as the total time of processing. All the results are given using MATLAB 2015a on a single thread with an Intel Core i5 CPU at 3.40GHz.

The speed of computation is summed up in table 3. The evaluation is done for the reconstruction of the same point set at different resolutions. The

Table 3: Comparison of computation speed compared to the resolution

Image size	600x550px	2400x2200px
Image resolution	1px = 4cm^2	1px = 1cm^2
Percentage of stripe holes	13%	61%
Percentage of occlusion holes	22%	25%
2D Projection	2.13s	3.78s
Diffusion	1.54s	3.27s
Mask extraction	0.18s	0.91s
Examplar-based inpainting	23.81s	6.31min
Total	**27.66s**	**6min38s**

choice of resolution and the amount of stripe holes do not affect much the computation time in proportion. However, the inpainting of large occlusions drastically increases the time of computation in the case of very high resolution. The computation speed of this step might be largely improved by using approaches derivated from Patch-Match (4). Moreover, the framework can be run in parallel as each step is independent of the next ones.

7. Conclusion and future work

We have proposed a complete framework to reconstruct high quality ground orthoimage from a point cloud aquired with LiDAR. This framework consists of several steps, which make use of classical modern imaging techniques. By taking into account the multi-modal nature of the data, we propose several modifications of these methods, leading to significantly better results.

The framework is designed to work automatically with a set of parameters that ensures satisfying results on a large variety of input data as demonstrated by the results. Our approach performs at least as well as previous techniques. In case of large occlusions or complex textures, it drastically outperforms earlier works in terms of visual quality. Moreover, robustness towards edges and structures conservation in both reflectance and height domain has been demonstrated.

Although the average results of the method are more than acceptable, it can underperform in some specific cases. Indeed, the recognition of the area to reconstruct using the envelop can sometimes fail when a massive non-static object in the scene is considered (*i.e. a building*), this area therefore not being reconstructed. Moreover, the use of examplar-based inpainting

introduces common issues such as aberrant synthesis when no similar patches are available.

In the future, we will focus on improving the current framework to better distinguish static structures and mobile objects, based on the work presented in (38). Moreover, we also want to perform labelling on the scene in order to produce a more relevant metric for evaluating the results. Finally, we aim at using aerial optical data in order to provide colored orthoimages, which promise very interesting challenges.

8. Acknowledgement

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 is a member of Institut Universitaire de France.

References

[1] G. Andreas, L. Philip, S. Christoph, and U. Raquel. Vision meets robotics: The KITTI dataset. *International Journal of Robotics Research*, 2013.

[2] G. Aubert and P. Kornprobst. *Mathematical problems in image processing: partial differential equations and the calculus of variations*. Springer, 2006.

[3] M.-F. Auclair-Fortier and D. Ziou. A global approach for solving evolutive heat transfer for image denoising and inpainting. *IEEE Trans. on Image Processing*, 15, 2006.

[4] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman. Patchmatch: A randomized correspondence algorithm for structural image editing. *ACM Trans. Graph.*, 28, 2009.

[5] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *ACM Comp. graphics and interactive techniques*, 2000.

[6] M. Bevilacqua, J.-F. Aujol, P. Biasutti, M. Brédif, and A. Bugeau. Joint inpainting of depth and reflectance with visibility estimation. *ISPRS Journal of the Photogrammetry, Remote Sens. and Spatial Inf. Sciences. (To appear)*, 2017.

[7] M. Bitenc, R. Lindenbergh, K. Khoshelham, and A. P. Van Waarden. Evaluation of a lidar land-based mobile mapping system for monitoring sandy coasts. *Remote Sensing*, 3, 2011.

[8] K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM Journal on Imaging Sciences*, 3, 2010.

[9] M. Brédif, B. Vallet, and B. Ferrand. Distributed dimensionality-based rendering of LIDAR point clouds. *ISPRS Arch. of the Photogrammetry, Remote Sens. and Spatial Inf. Sciences*, 40, 2015.

[10] J. E. Bresenham. Algorithm for computer control of a digital plotter. *IBM Systems journal*, 4, 1965.

[11] P. Buyssens, M. Daisy, D. Tschumperlé, and O. Lézoray. Exemplar-based inpainting: Technical review and new heuristics for better geometric reconstructions. *IEEE Trans. on Image Processing*, 24, 2015.

[12] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Jour. of Math. Imag. and Vis.*, 40(1), 2011.

[13] H. Chen, M. Cheng, J. Li, and Y. Liu. An iterative terrain recovery approach to automated DTM generation from airborne LiDAR point clouds. *ISPRS Arch. of the Photogrammetry, Remote Sens. and Spatial Inf. Sciences*, 39, 2012.

[14] Q. Chen, H. Wang, H. Zhang, M. Sun, and X. Liu. A point cloud filtering approach to generating DTMs for steep mountainous areas and adjacent residential areas. *Remote Sensing*, 8, 2016.

[15] Z. Chen, B. Gao, and B. Devereux. State-of-the-art: DTM generation using airborne LIDAR data. *Sensors*, 17, 2017.

[16] A. Criminisi, P. Pérez, and K. Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE Trans. on Image Processing*, 13, 2004.

[17] I. Daribo and B. Pesquet-Popescu. Depth-aided image inpainting for novel view synthesis. In *IEEE Trans. on Image Processing*, 2010.

[18] A. A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, 1999.

[19] S. El-Halawany, A. Moussa, D. Lichti, and N. El-Sheimy. Detection of road curb from mobile terrestrial laser scanner point cloud. In *ISPRS Arch. of the Photogrammetry, Remote Sens. and Spatial Inf. Sciences*, volume 2931, 2011.

[20] H. Guan, J. Li, Y. Yu, L. Zhong, and Z. Ji. DEM generation from lidar data in wooded mountain areas by cross-section-plane analysis. *International Journal of Remote Sensing*, 35, 2014.

[21] T. Hackel, N. Savinov, L. Ladicky, J. D. Wegner, K. Schindler, and M. Pollefeys. Semantic3D.net: A new large-scale point cloud classification benchmark. *ISPRS Arch. of the Photogrammetry, Remote Sens. and Spatial Inf. Sciences*, 2017.

[22] A. Hervieu and B. Soheilian. Road side detection and reconstruction using LIDAR sensor. In *IEEE Intelligent Vehicles Symposium*, volume 1, 2013.

[23] A. Hervieu and B. Soheilian. Semi-automatic road/pavement modeling using mobile laser scanning. *ISPRS Arch. of the Photogrammetry, Remote Sens. and Spatial Inf. Sciences*, 2, 2013.

[24] A. Hervieu, B. Soheilian, and M. Brédif. Road marking extraction using a model&data-driven RJ-MCMC. *ISPRS Arch. of the Photogrammetry, Remote Sens. and Spatial Inf. Sciences*, 2, 2015.

[25] X. Hu, L. Ye, S. Pang, and J. Shan. Semi-global filtering of airborne LiDAR data for fast extraction of digital terrain models. *Remote Sensing*, 7, 2015.

[26] J. J. Koenderink. The structure of images. *Biological cybernetics*, 50, 1984.

[27] K. Kraus and N. Pfeifer. Advanced DTM generation from LIDAR data. *ISPRS Arch. of the Photogrammetry, Remote Sens. and Spatial Inf. Sciences*, 34, 2001.

[28] L. Lorenzi, F. Melgani, and G. Mercier. Inpainting strategies for reconstruction of missing data in vhr images. *IEEE Trans. on Geoscience and Remote Sensing Letters*, 8, 2011.

[29] C.-P. McElhinney, P. Kumar, C. Cahalane, and T. McCarthy. Initial results from European Road Safety Inspection (EURSI) mobile mapping project. In *ISPRS Arch. of the Photogrammetry, Remote Sens. and Spatial Inf. Sciences*, volume 38, 2010.

[30] X. Meng, N. Currit, and K. Zhao. Ground filtering algorithms for airborne LiDAR data: A review of critical issues. *Remote Sensing*, 2, 2010.

[31] P. Pérez, M. Gangnet, and A. Blake. Poisson image editing. In *ACM Trans. on Graphics*, volume 22, 2003.

[32] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12, 1990.

[33] S. Pu, M. Rutzinger, G. Vosselman, and S. O. Elberink. Recognizing basic structures from mobile laser scanning data for road inventory studies. *ISPRS Arch. of the Photogrammetry, Remote Sens. and Spatial Inf. Sciences*, 66, 2011.

[34] X. Qu, B. Soheilian, and N. Paparoditis. Vehicle localization using mono-camera and geo-referenced traffic signs. In *IEEE Intelligent Vehicles Symposium*, volume 4, 2015.

[35] J. Rabin, G. Peyré, J. Delon, and M. Bernot. Wasserstein barycenter and its application to texture mixing. In *Scale Space and Variational Methods in Computer Vision*. Springer, 2011.

[36] F. Rottensteiner and C. Briese. A new method for building extraction in urban areas from high-resolution LIDAR data. *ISPRS Arch. of the Photogrammetry, Remote Sens. and Spatial Inf. Sciences*, 34, 2002.

[37] M. Rubinstein, A. Shamir, and S. Avidan. Improved seam carving for video retargeting. In *ACM Trans. on Graphics*, volume 27, 2008.

[38] J. R. Schoenberg, A. Nathan, and M. Campbell. Segmentation of dense range information in complex urban scenes. In *IEEE Trans. on Intelligent Robots and Systems*, 2010.

[39] A. Serna and B. Marcotegui. Urban accessibility diagnosis from mobile laser scanning data. *ISPRS Arch. of the Photogrammetry, Remote Sens. and Spatial Inf. Sciences*, 84, 2013.

[40] J. Serra. *Image analysis and mathematical morphology*, volume 1. Academic Press, 1982.

[41] J. Shan and C. K. Toth. *Topographic laser ranging and scanning: principles and processing.* CRC Press, 2008.

[42] Y.-C. Shao and L.-C. Chen. Automated searching of ground points from airborne lidar data using a climbing and sliding method. *Photogrammetric Engineering & Remote Sensing*, 74, 2008.

[43] O. Tournaire, B. Soheilian, and N. Paparoditis. Towards a sub-decimetric georeferencing of groundbased mobile mapping systems in urban areas: Matching ground-based and aerial-based imagery using roadmarks. *ISPRS Arch. of the Photogrammetry, Remote Sens. and Spatial Inf. Sciences*, 36, 2006.

[44] B. Vallet and J.-P. Papelard. Road orthophoto/DTM generation from mobile laser scanning. *ISPRS Arch. of the Photogrammetry, Remote Sens. and Spatial Inf. Sciences*, 3, 2015.

[45] R. Wack and A. Wimmer. Digital terrain models from airborne laserscanner data-a grid based approach. *ISPRS Arch. of the Photogrammetry, Remote Sens. and Spatial Inf. Sciences*, 34, 2002.

[46] L. Wang, H. Jin, R. Yang, and M. Gong. Stereoscopic inpainting: Joint color and depth completion from stereo images. In *IEEE Conf. on Computer Vision and Pattern Recognition*, 2008.

[47] J. Weickert. *Anisotropic diffusion in image processing*, volume 1. Teubner Stuttgart, 1998.

[48] W. Xiaobo, W. Shixing, and X. Chunsheng. A new study of delaunay triangulation creation. *ACTA Geodaetica et Cartographic Sinica*, 1, 1999.

[49] K. Zakšek and N. Pfeifer. An improved morphological filter for selecting relief points from a LIDAR point cloud in steep areas with dense vegetation. *Technical report: Institute of Anthropological and Spatial Studies, Ljubljana, Slovenia*, 2006.

[50] J. Zhang and X. Lin. Filtering airborne LiDAR data by embedding smoothness-constrained segmentation in progressive TIN densification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 81, 2013.

[51] G. Zhao and J. Yuan. Curb detection and tracking using 3D-LIDAR scanner. In *IEEE Int. Conf. on Image Processing*, volume 19, 2012.

[52] L. Zhuang and J. M. Bioucas-Dias. Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11, 2018.