Diffusion and inpainting of reflectance and height LiDAR orthoimages

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

▶ **To cite this version:**

Pierre Biasutti, Jean-François Aujol, Mathieu Brédif, Aurélie Bugeau. Diffusion and inpainting of reflectance and height LiDAR orthoimages. 2016. hal-01322822v1

HAL Id: hal-01322822

https://hal.science/hal-01322822v1

Preprint submitted on 27 May 2016 (v1), last revised 12 Feb 2019 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Diffusion and inpainting of reflectance and height LiDAR orthoimages

P. Biasutti[123] J-F. Aujol[1] M. Brédif[3] A. Bugeau[2]
[1] IMB, CNRS, IPB, University of Bordeaux
[2] LaBRI, CNRS, University of Bordeaux
[3] Université Paris-Est, IGN, SRIG, MATIS, 73 avenue de Paris, 94160 Saint Mandé, France
{pierre.biasutti, jean-francois.aujol}@math.u-bordeaux.fr,
mathieu.bredif@ign.fr,
aurelie.bugeau@labri.fr

Abstract

This paper presents a fully automatic framework for the production of both reflectance and height orthoimages at very high resolution (1cm^2 per pixel) from a mobile laser scan. Traditional orthoimages are obtained using aerial techniques which are limited by their accuracy and the occlusions due to hovering objects (trees, tunnels, bridges ...). Mobile laser scan is an acquisition technique made at ground level which overcomes those limitations.

In this paper, we propose a framework to reconstruct high quality ground orthoimages from a point cloud acquired via mobile laser scan in a urban environment.

The point cloud provided by the laser scan is first projected on a 2D-pixel grid to generate sparse and noisy reflectance and height images. Dense images are computed using an anisotropic diffusion algorithm taking both channels into account. Next, we extract a mask of the large regions where diffusion does not provide a visually satisfying result. Finally, we propose an examplar-based inpainting method to fill-in these regions. It relies on both channels and an assumption about the alignment between structures to inpaint.

A number of examples on real acquisition data demonstrates the effectiveness of the proposed pipeline in generating a very high resolution orthoimage of reflectance and height while filling holes of various sizes in a visually satisfying way.

1 Introduction

Orthophotography and Digital Terrain Model (DTM), defined as respectively 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, environmental studies... These techniques are often computed from aerial acquisition devices such as satellites, aircrafts and more recently unmanned aerial vehicles (UAVs). Although these techniques provide a very well known and common approach to the problem of orthoimage generation, they are limited in terms of accuracy and resolutions and suffer from occlusions that are caused by the natural and urban environnement such as trees, tunnels, tall buildings, ...

In the past decade, a new technology known as mobile mapping has emerged. It provides a new acquisition context in which the data is acquired at ground level from a moving vehicle. Among the variety of mobile mapping techniques, we focus on the laser detection and ranging (LiDAR) acquisition technology : the mobile laser scan (MLS). The proximity of the acquisition ensures a high resolution as well as a diminution of occluded areas. Thus, we advocate that orthoimages of both reflectance and ground height can be produced at very high resolution (1cm^2 per pixel) using MLS to overcome the previously mentioned limitations of traditional techniques.

1.1 Laser detection and ranging

LiDAR is a technique used to create a 3D point cloud of an object, a room or even an open area such as urban environment, fields... A sensor is used to illuminate a target by emitting a laser beam and measuring the position of the target point and the reflectance of the beam. The operation is repeated many times while variating the angle of emission and by shifting the sensor in order to create a very accurate (subcentimetric) 3D mapping of the target object. Typically, the sensor is put on a vehicle following the road at a relatively constant speed. The resulting point cloud provides several components such as the 3D coordinates in both earth and the sensor referentials, the distance between the measured point and the sensor (range) and the reflectance in decibels. The reflectance is estimated from the backscattered energy as the albedo of the target, assuming a fronto-parallel Lambertian target surface. Therefore, it does not provide information that is directly related to a color information in the visible domain but it can be effectively used as a single channel texture information for the production of an orthoimage.

1.2 High resolution orthoimaging

High resolution orthoimaging for reflectance and height has many possible applications in a urban context such as public building accessibility for soft mobilities (disabled, strollers)

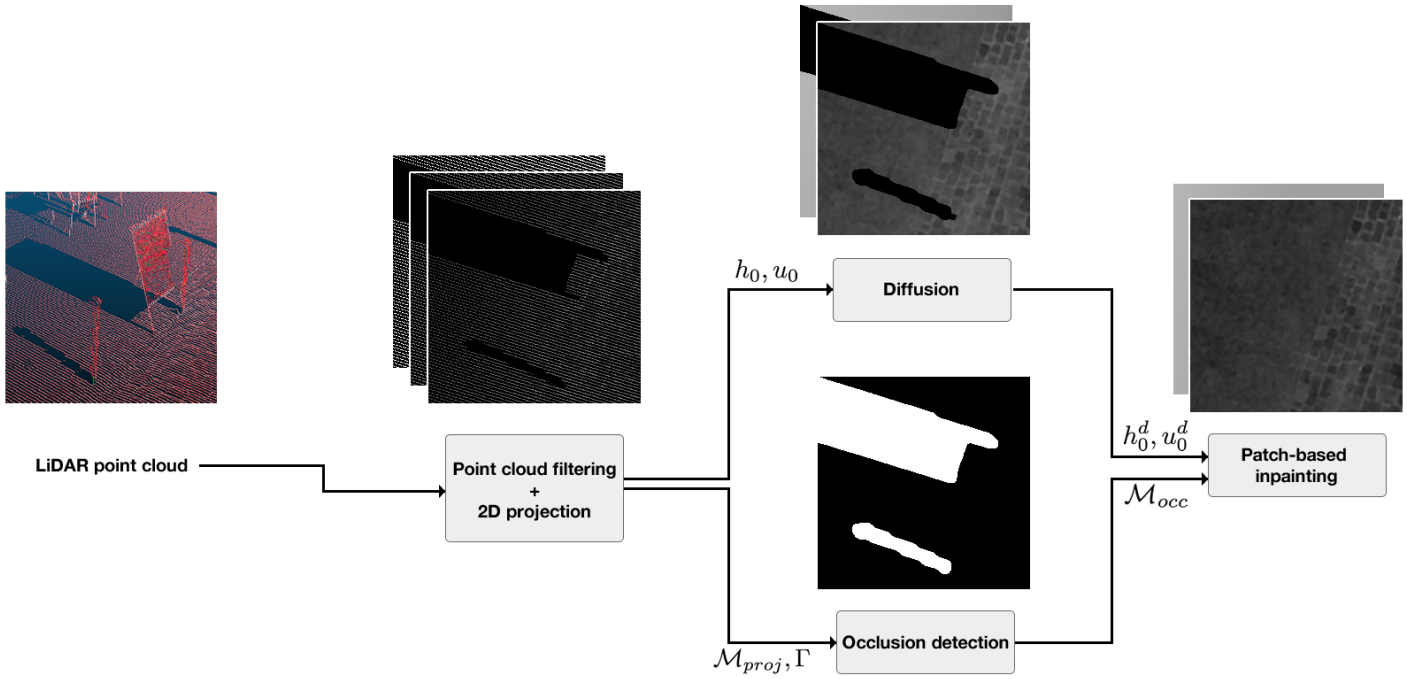


Figure 1: Full orthoimage production pipeline from MLS. Framed rectangles are processing steps, arrows are exchanged data.

and itinerary computations [25], precise mapping of road marks [12], road limits or curbs [15, 9, 31, 11], road inventory [18], road surface modelling and quality measurements [10], mobile mapping registrations on aerial images [27], image based localization using ground landmarks [19], etc.

Moreover, some recent legislations in European countries call for a subdecimetric accuracy mapping of buried objects (water and gaz pipes, internet and phone wires, etc.) as the lack of accurate data has lead to accidents and delays in many public works. Very high resolution orthoimaging could help in meeting the requirements of those legislations.

The projection of a MLS point cloud at centrimetric resolutions creates a sparse image due to its inhomogeneous density. The problem of filling in the holes created by the lack of information has already been explored in [28]. The authors perform Poisson interpolation [16] in order to fill the projection. However, in case of large holes, the reconstruction is not visually satisfying. Moreover, the interpolation fills the whole image instead of limiting the reconstruction to the ground.

1.3 Problem statement and 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 contributions are to provide a fully automatic framework with an efficient set of default parameters. Moreover, we introduce a novel anisotropic diffusion technique that uses both texture and height information as well as an examplar-based inpainting method that is modified in order to take advantage of all the available data provided by the LiDAR sensor.

1.4 Outline of the paper

In this paper, we propose a novel approach for the reconstruction of ground orthoimages that produces very satisfying results in terms of visual quality and coherence (2). We first explain how the point cloud can be projected on a 2D-pixel grid after filtering ground points (3). 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 (4). We show 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 (5). Finally, in 6, we validate our framework by presenting various examples of reconstruction made on real acquisition data.

2 Framework description

Orthoimage generation from MLS is often done by a successive combination of algorithms. In this paper, we aim at presenting an efficient and fully automatic pipeline to reconstruct an aerial image from a LiDAR point cloud. The proposed framework is summed up in figure 1.

From the point cloud, we need to extract the points that correspond to the ground only by computing an envelop Γ (see section 3) and we project both reflectances and height information in two images: u_0 and h_0. We also build a mask $\mathcal{M}_{proj}$ of the pixels where at least one point was projected.

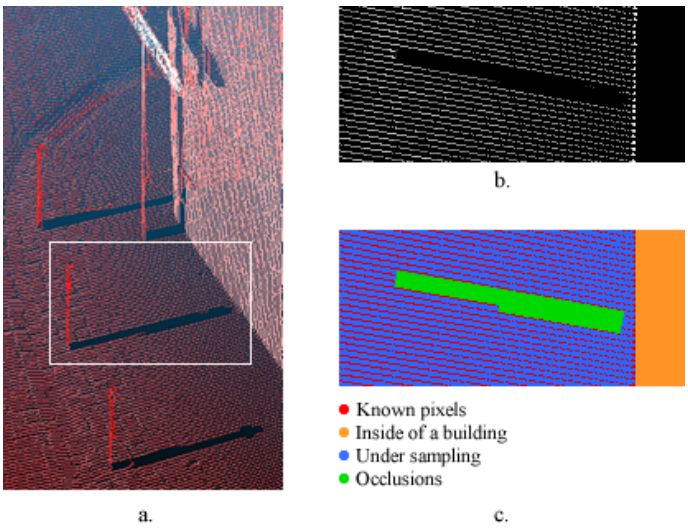


Figure 2: Highlighting of the different types of holes in the sparse projection. (a) is the original point cloud colored with the reflectance, (b) is the sparse projection of the red rectangle in (a) after filtering, (c) is the sparse projection labelled with the different kinds of holes.

At this point, the projections u_0 and h_0 are sparse as they do not cover all the pixels in the images. We can distinguish 3 classes of holes for which the images are sparse:

- Areas that do not correspond to the ground points (inside of buildings, point that are too far from the sensor, ...) which can easily be detected using Γ. Those areas do not require any more processing as they are not part of the reconstruction.
- Under sampling of the LiDAR sensing which may be caused by scan line separation distances (exagerated by high vehicle speed) or within a scan line at large distances.
- Occlusions that appear when a relatively vertical object was blocking the beams and prevents the sensor from acquiring ground points.

Figure 2 sums up the different kinds of holes that can appear in the projection. Some parts of the projection correspond to the inside of a building (figure 2.c in orange), under sampling holes appear in between lines of acquisition (figure 2.c in blue) and an occlusion is caused by a pole blocking the laser beams (figure 2.c 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 dense images are respectively called u_0^d and h_0^d. The occlusion mask $\mathcal{M}_{occ}$ is then retrieved using the projection mask and mathematical 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 tideous as we have to filter groups of point that represent relatively planar structures. Horizontal planes may usually be considered as ground structures. However, modern MLS tends 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 [22] or by performing advanced structural analysis [14, 3]. However, these solutions have shown their limitations in the case of a scenario where a huge variety of different structures are present such as a urban environment.

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. We first create an empty envelop of the size of the projection where each pixel has an infinite value. 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 [4]. 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. Then, we apply a threshold to limit too high points compared to the ground level under the aquisition device. Figure 3 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 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. Using the median of all the values of the points projected in the same pixel instead of the mean value could have been a more natural choice to prevent noise from altering the

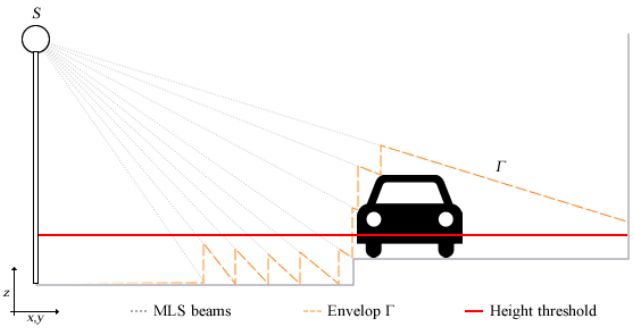


Figure 3: Slice of an envelop Γ obtained by evaluating several aligned beams coming from the sensor S until they hit an object of the scene. The red line is the final threshold applied to the envelop to exclude too high points.

result. However in our tests the amount of points projected in every pixel as well as the proximity between them ensures a relatively homogeneous result. Moreover, the computation of the mean is more cost efficient than the one of the median.

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 is arbitrary but in the case of a urban scenario, the ground components are likely to be under 1m above the road.

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 [5], the generalized total variation [2], structure tensor diffusion [30] or partial differential equation diffusion [1].

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 [13]. 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 [17] algorithm is a well known algorithm for anisotropic diffusion. It is partially inspired from the Gaussian diffusion and is defined as follows:

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

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 keeping the computation of homogenous transitions between sampled scan lines. Nevertheless, this technique only takes into account the gradients of one channel. In our current 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:

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

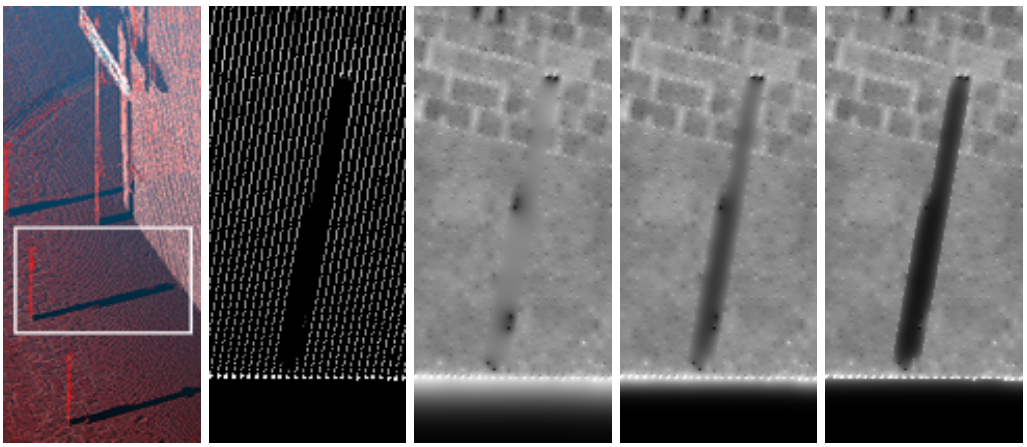
where we recall that u_0 is the reflectance image and h_0 is the height image. We introduce the new weighting function f that emerges from the one used in equation (1) as follows:

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

having α, β as weighting constants quantifying how gradients of reflectance and height need to be considered. Using this modification, we can now take into account gradients coming from both u_0 and h_0.

4.3 Comparison with other diffusion techniques

Figure 4 illustrates a comparison between different diffusion methods. We can see in figure 4.c that the Gaussian diffusion does not preserve any edges. The Perona-Malik algorithm gives a better result but the diffusion is not correctly



a b c d e

Figure 4: 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.

Table 1: Statistical properties of diffusion methods (computed on reflectances)

Diffusion technique	Mean value	STD
Ground truth	**89.180**	**1.349**
Gaussian	88.167	0.577
Anisotropic [17]	88.996	1.223
Anisotropic + altitudes	**89.056**	**1.326**

blocked at the border of large holes (figure 4.d). This can create discontinuities and create artifacts in the examplar-based inpainting performed in the next section. Our proposed algorithm succeeds in filling the holes due to undersampling and correctly blocks the diffusion at the borders of high variation of height (occlusions, inside of buildings.)

Another important aspect of our modification is the preservation of the second order statistical properties. The noise is kept at the same level in reconstructed areas as it is in known areas, and the result is suitable for being used in a statistical study. Table 1 presents the results of the computation of the mean and the standard deviation of the reconstructed areas using different kinds of diffusions. The comparison is made on 10^4 px by comparing only the reconstruction of undersampling holes against the ground truth (i.e. area where the aquisition device evolves at low speed which prevents undersampling holes from appearing). The proposed algorithm provides the closest results to the ground truth.

4.4 Parameters

The number of iterations has to be chosen in order to fill in stripe holes. It depends on the chosen resolution. A good speed-up can be obtained by first performing a nearest neighbor interpolation on both u_0 and h_0 independently as it lowers drastically the number of iterations needed. 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. 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 ...) 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 [26] 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.

As explained above, occlusion holes and stripe holes are not the only kind of missing data after the projection. We need to remove the pixels that are out of the final reconstruction (inside of buildings, regions that are too far from the sensor ...). Typically, those regions correspond to the parts where the envelop is not defined in the image and can easily be removed from the mask obtained after the closing.

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 [8] or the full patch [6]. The operation is repeated until the hole is fully closed. Novel approaches allow to reconstruct the texture using both color information and depth information such as [7] and [29]. However those 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. [6] algorithm that was designed for the well preservation of the structures in the reconstruction. In [6], 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 [6] 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 SSD of the candidate patch in both the reflectance and the height images. For each candidate, a score is attributed by combining both channels as follows:

$$S_p(P_t, P_c) = \mathrm{SSD}(\mathrm{P_t^R}, \mathrm{P_c^R}) + \eta \times \mathrm{SSD}(\mathrm{P_t^H}, \mathrm{P_c^H}) \qquad (4)$$

where P_t is the target patch to be filled, P_c is a candidate patch, 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. We recall that SSD is the sum of squared differences defined as follows:

$$\mathrm{SSD}(\mathrm{P_1}, \mathrm{P_2}) = \sum_{\mathrm{i,j} \in \Omega} (\mathrm{P_1(i,j)} - \mathrm{P_2(i,j)})^2 \qquad (5)$$

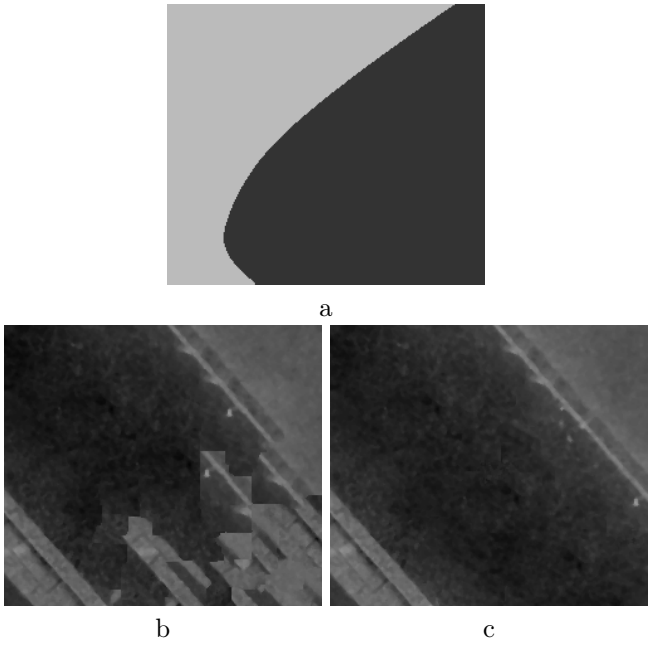


Figure 5: Comparison of [6] and our proposed modification on the junction between the road and a pavement. (a) is the mask where the darkest region is being reconstructed using examplar-based inpainting, (b) the result from [6], (c) our proposed optimization. The result is clearly better in (c) as the reconstruction conserves the structures of the image without creating new artifacts such as the one appearing on the left of (b).

The impact of the use of the height map in the synthesis is very noticeable in figure 5. 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). 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 6. Therefore, we can constrain

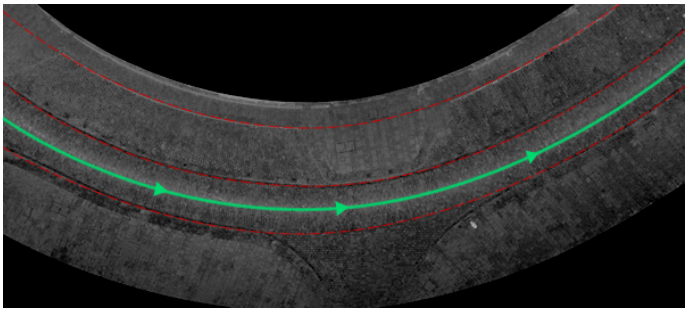


Figure 6: Illustration of the assumption that the urban environment evolves in a similar way than the path of the sensor. The straight green line shows the path of the sensor. Each dashed line represents areas of same distance to the sensor.

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:

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

having $d(P)$ the distance between the sensor and the center of the patch P and γ a regularization parameter that allows to constrain 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 [23].

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 those 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 application of this study was to be able to provide very high quality orthoimages. Therefore, all reconstructions were done at the maximal resolution allowed by 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.

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 comparison between traditional aerial orthophotography and MLS orthoimage using our framework is given in figure 7. The resolution provided by the 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.

In figure 8, we show a visual comparison between the proposed framework and the method introduced in [28]. 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.

Various examples of how good reconstructions are using our method are displayed in figure 9. We can see on figure 9.a that the framework performs a very good reconstruction of fine details such as cobbles. In figure 9.b, 25% of the image is occluded ($\sim 5.10^5$px), mostly because of cars. This demonstrates how the coupling of reflectances, heights and

Figure 7: Comparison between aerial orthophotography (top) and MLS orthoimage (bottom). The MLS reconstruction contains $\sim 6.10^6$px in which 72% were obtained by diffusion, and 10% by examplar-based inpainting.

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**
Vallet et al. [28]	1.87	0.78	2.05	0.80

ranges in the examplar-based inpainting method performs well on large scale reconstruction. Finally, figure 9.c shows an extreme scenario where the use of the range is relevant. The environment is reconstructed (16%, $\sim 10^6$px) while preserving the structure of the road.

The purpose of this pipeline is to generate both reflectance and height orhtoimages. In figure 10, we show how the two outputs can be combined in order to obtain a 3D model of the road. Figure 10.a and 10.b are the reflectance image and the height image of the area that is being modelled in figure 10.c. We can see that the 3D model respects the topography of the scene with the junction of the road and a pavement.

6.3 Quantitative analysis

Appart from the visual results, we also provide a numerical comparison between the proposed framework and the one of [28]. 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, correlation, ...). In the case of texture synthesis, the similarity cannot be 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 measure of the standard deviation and the distance between histogram, also known as Wasserstein metric in [21, 20], provide simple and efficient metrics for evaluating the quality of our results.

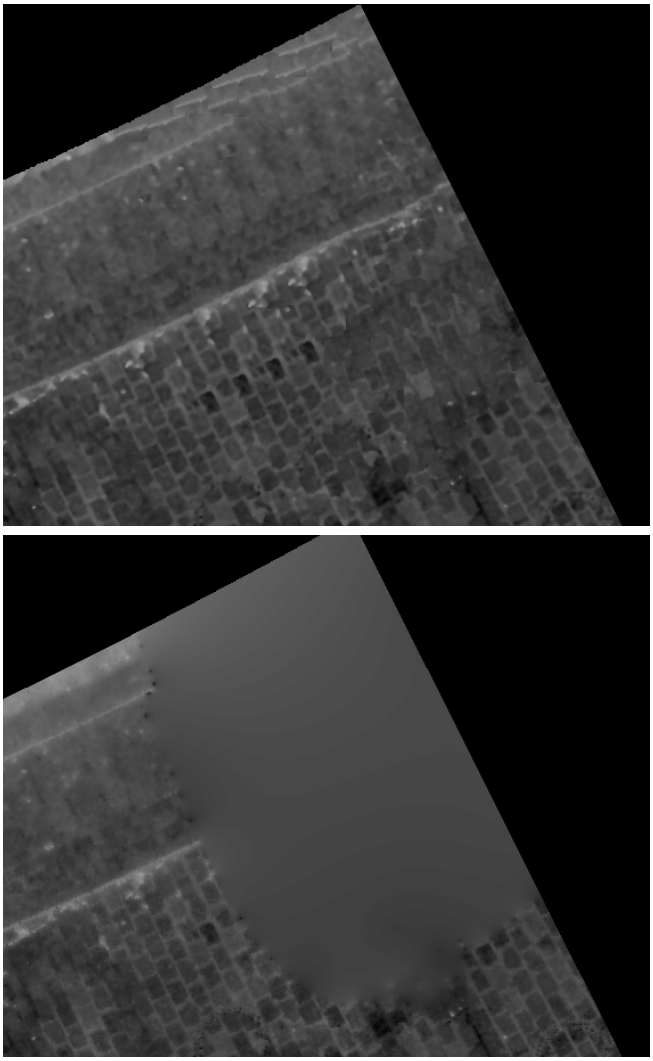


Figure 8: Comparison between our proposed framework (top) and the one introduced in [28] (bottom). Texture is better preserved using our framework.

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 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 11. 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

Figure 9: Various results on different types of textures. (a) reconstruction of large areas containing cobbles, (b) reconstruction of straight road and (c) reconstruction in a curved path.

below 1cm. This validates the proposed framework for the reconstruction of height map.

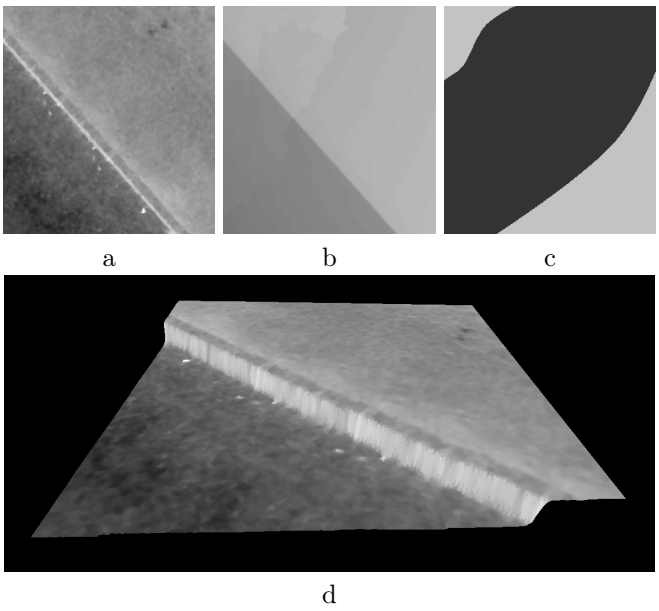


Figure 10: 3D model of the ground of a part of an orthoimage. (a) is the reflectance image used as texture for the 3D model, (b) is the height image used as the height coordinate of the 3D model, (c) is the mask where the darkest region was reconstructed using examplar-based inpainting, (d) is the 3D model obtained using both reflectance and height orthoimages.

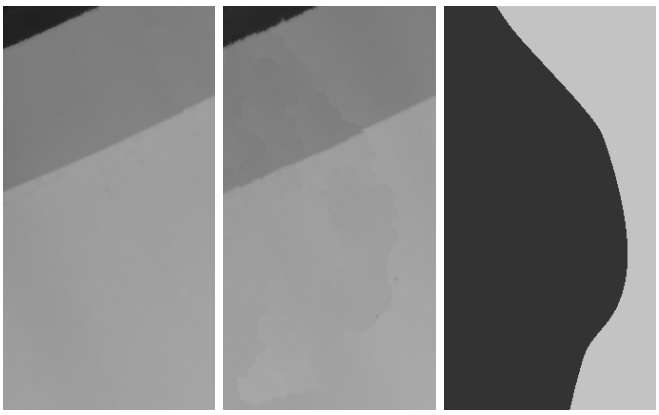


Figure 11: Comparison between height images with and without occlusion on the junction between a road and a pavement. (a) is the original height map, (b) is the reconstruction of an occlusion in the same area. The occlusion corresponds to the darkest region of (c). The mean square error of the reconstruction (b) compared to (a) on the occlusion region is 2mm.

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

Table 3: Computation speed comparison

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	6.31m
Total	**27.66s**	**6m38s**

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 cloud of points at different resolutions. The choice of resolution and the amount of stripe holes do not affect much the computation time. However, the inpainting of large occlusions drastically increases the time of computation in the case of very high resolution. Those results could efficiently be lowered by using GPU computing. 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 a building and therefore is not 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 [24]. 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 Aknowledgement

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

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

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

[4] J. E. Bresenham. Algorithm for computer control of a digital plotter. *IBM Systems journal*, 4(1):25–30, 1965.

[5] 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.

[6] A. Criminisi, P. Pérez, and K. Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE Trans. on Image Processing*, 13(9), 2004.

[7] I. Daribo and B. Pesquet-Popescu. Depth-aided image inpainting for novel view synthesis. In *IEEE International Workshop on Multimedia Signal Processing*, 2010.

[8] 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.

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

[10] A. Hervieu and B. Soheilian. Road side detection and reconstruction using lidar sensor. In *IEEE Intelligent Vehicles Symposium*, volume 1, pages 1247–1252, 2013.

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

[12] A. Hervieu, B. Soheilian, and M. Brédif. Road marking extraction using a model&data-driven rj-mcmc. *Int. Arch. of the Photogrammetry, Remote Sens. and Spatial Inf. Sciences*, 2(3), 2015.

[13] Jan J Koenderink. The structure of images. *Biological cybernetics*, 50(5), 1984.

[14] K. Kraus and N. Pfeifer. Advanced dtm generation from lidar data. *Int. Arch. of the Photogrammetry, Remote Sens. and Spatial Inf. Sciences*, 34(3/W4), 2001.

[15] C-P. McElhinney, P. Kumar, C. Cahalane, and T. McCarthy. Initial results from european road safety inspection (eursi) mobile mapping project. In *Int. Arch. of the Photogrammetry, Remote Sens. and Spatial Inf. Sciences*, volume 38, pages 440–445, 2010.

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

[17] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(7), 1990.

[18] S. Pu, M. Rutzinger, G. Vosselman, and S. O. Elberink. Recognizing basic structures from mobile laser scanning data for road inventory studies. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(6), 2011.

[19] X. Qu, B. Soheilian, and N. Paparoditis. Vehicle localization using mono-camera and geo-referenced traffic signs. In *IEEE Intelligent Vehicles Symposium*, volume 4, pages 605–610, 2015.

[20] J. Rabin and G. Peyré. Wasserstein regularization of imaging problem. In *IEEE Int. Conf. on Image Processing*, pages 1541–1544, 2011.

[21] 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*, pages 435–446. 2011.

[22] F. Rottensteiner and C. Briese. A new method for building extraction in urban areas from high-resolution lidar data. *Int. Arch. of the Photogrammetry, Remote Sens. and Spatial Inf. Sciences*, 34(3/A):295–301, 2002.

[23] M. Rubinstein, A. Shamir, and S. Avidan. Improved seam carving for video retargeting. *ACM Trans. on Graphics*, 27(3), 2008.

[24] J. R Schoenberg, A. Nathan, and M. Campbell. Segmentation of dense range information in complex urban scenes. In *IEEE International Conference on Intelligent Robots and Systems (IROS)*, pages 2033–2038, 2010.

[25] A. Serna and B. Marcotegui. Urban accessibility diagnosis from mobile laser scanning data. *Int. Arch. of the Photogrammetry, Remote Sens. and Spatial Inf. Sciences*, 84:23–32, 2013.

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

[27] 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. *Int. Arch. of the Photogrammetry, Remote Sens. and Spatial Inf. Sciences*, 36, 2006.

[28] B. Vallet and J. P. Papelard. Road orthophoto/DTM generation from mobile laser scanning. *Int. Arch. of the Photogrammetry, Remote Sens. and Spatial Inf. Sciences*, 2, 2015.

[29] 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.

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

[31] G. Zhao and J. Yuan. Curb detection and tracking using 3d-lidar scanner. In *IEEE Int. Conf. on Image Processing*, volume 19, pages 437–440, 2012.