HAL
open science

Fast Image and LiDAR alignment based on 3D rendering in sensor topology

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

To cite this version:

Pierre Biasutti, Jean-François Aujol, Mathieu Brédif, Aurélie Bugeau. Fast Image and LiDAR alignment based on 3D rendering in sensor topology. 2019. hal-02100715

HAL Id: hal-02100715

https://hal.science/hal-02100715

Preprint submitted on 16 Apr 2019

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.

Fast Image and LiDAR alignment based on 3D rendering in sensor topology

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

[a] *Univ. Bordeaux, LaBRI, INP, CNRS, UMR 5800, F-33400 Talence, France*
[b] *Univ. Bordeaux, IMB, INP, CNRS, UMR 5251, F-33400 Talence, France*
[c] *Univ. Paris-Est, LASTIG GEOVIS, IGN, ENSG, F-94160 Saint-Mandé, France*

Abstract

Mobile Mapping Systems are now commonly used in large urban acquisition campaigns. They are often equiped with LiDAR sensors and optical cameras, providing very large multimodal datasets. The fusion of both modalities serves different purposes such as point cloud colorization, geometry enhancement or object detection. However, this fusion task cannot be done directly as both modalities are only coarsely registered. This paper presents a fully automatic approach for LiDAR projection and optical image registration refinement based on LiDAR point cloud 3D renderings. First, a coarse 3D mesh is generated from the LiDAR point cloud using the sensor topology. Then, the mesh is rendered in the image domain. After that, a variational approach is used to align the rendering with the optical image. This method achieves high quality results while performing in very low computational time.

Results on real data demonstrate the efficiency of the model for aligning LiDAR projections and optical images.

1. Introduction

Over the past decades, the interest in urban acquisition systems has been growing continuously, especially Mobile Mapping Systems (MMS). These systems are vehicles equipped with many sensors that produce different modalities in order to acquire all the details of a scene. These systems are used on wide acquisition campaigns in cities, roads, highways, resulting in the production of very large - multimodal - datasets. Among all available sensors often met on such systems, 3D LiDAR sensors joint with optical cameras enable a geometrical acquisition of the scene as well as the acquisition of textures and colors. However, due to the complexity of such acquisition systems, the calibration from one sensor to the other does not generally meet pixel accuracy. This can be caused by the instability of the sensors throughout a mobile acquisition, where the calibration slowly deteriorates while the systems are being operated. Therefore, the different modalities are slightly misaligned which can compromise further processing requiring data fusion, *e.g.* point cloud colorization or multimodal object detection which may be required for autonomous driving applications or to filter out non-permanent objects for cartographic purposes. It is possible to interactively reduce this misalignment, but most of the time this solution is not suitable as the datasets are typically composed of thousands of examples. The automatic alignement of LiDAR data to optical image is therefore a crucial issue.

The problem of LiDAR to image alignment rises several issues. First of all, the comparison between the two modalities can only be done if they share common attributes (colors or reflectances). However, in many systems, each sensor solely acquires a specific aspect of the scene. Moreover, optical sensors and LiDAR sensors are located at different positions on the MMS. This implies that the different sensors do not acquire the scene from the same point of view, resulting in visual ambiguities. The correlation between both modalities is therefore irrelevant for some parts of each data.

The paper contribution is two-fold, as illustrated Figure 1: first, we propose a very fast approach for mesh generation from 3D LiDAR data using sensor topology. The second contribution of the paper is an extension of a variational multimodal registration method (Sutour et al.) to the problem of LiDAR to image alignment with higher degree of freedom.

The paper is organized as follows: first, an overview of the related works is presented. Then, the mesh generation from the sensor topology is explained. After that, the extension of the variational method is detailed. Finally, evaluation and results are shown and a conclusion is drawn.

2. Previous works

Multi-modal registration has been a subject of interest over the past decades. In this section, previous works on multi-modal registration as well as previous works on LiDAR to optical image are introduced.

2.1. Multi-modal image registration

In computer vision, registration methods often consist in the detection and the matching of corresponding features from two different modalities to recover the 2D transformation that provides the best alignment between the two input images. It often assumes that the displacement between both image is small, otherwise there would not be any good alignment between the two images because of the differences in perspective. Feature points are extracted using common methods (SIFT (Lowe) or SURF (Bay et al., 2006)), or more specified adaptations (Mikolajczyk and Schmid; Rublee et al.). These features are then matched using the RANSAC algorithm (Fischler and Bolles) to estimate the optimal transformation, as it can be seen in many biomedical imaging works (Allaire et al., 2008; Paganelli et al., 2012; Toews et al., 2013). However, these methods rely on strong similarities

Email address: pierre.biasutti@labri.fr (Pierre Biasutti)

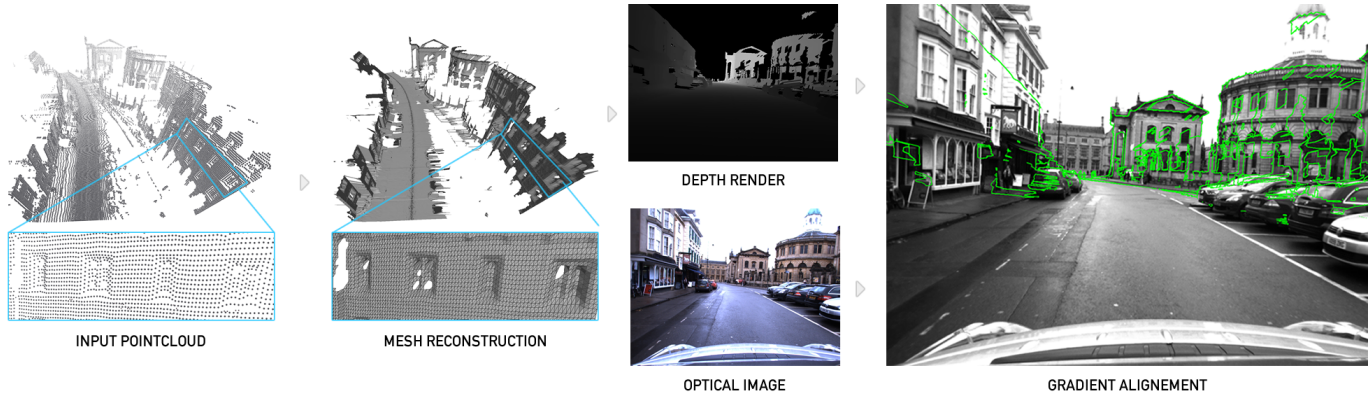

Figure 1: Scheme of the proposed framework.

between each modalities which can be limited in a multimodal context. This problem can also be solved using variational approaches. In this case, the optimal alignment can be defined as the maximum of a given metric, typically Mutual Information (Viola and Wells III) or Cross-correlation (Roshni and Revathy, 2008), which aim at finding correlations between two distributions of intensities. These methods perform well as long as there exists a bijection between both modalities (*e.g.* between CT and MR images) which is not the case between LiDAR range measurements and optical images. Another approach presented in (Sutour et al.) aligns the gradients of both modalities, thus being agnostic to any correlation between the modalities. However, this method only estimates translation and scaling without rotation. Moreover, it implies that both modalities are defined in the same domain and that gradients can be computed on both modalities, which is typically not possible when dealing with 3D points.

2.2. LiDAR to optical registration

The problem of LiDAR to optical registration can be divided into three main kinds of approaches: *2D feature-based methods*, *3D-based method* and *statistical methods*.

2D feature-based methods aim at establishing correspondences between feature points of the optical image and the point cloud projected in the optical image domain. In (Moussa et al.), the authors propose a method that uses ASIFT features (Morel and Yu) to match a colorized point cloud with an optical image. Aberrant correspondences are then filtered out using RANSAC (Fischler and Bolles). The final 3D pose is estimated by solving a Perspective-n-Point problem (Lepetit et al.) in which the 2D coordinates of feature points in the optical image is associated with the 3D locations of the corresponding feature points in the point cloud. González et al. propose a method for estimating the location of an optical image relatively to a 3D colorized point cloud of the same scene. The image is first enhanced to increase its contrasts. Then, the projection of the point cloud is manually resized in order to fit the optical image as well as possible. After that, correspondences are estimated by averaging cross-correlation and least square metrics. Finally, the 3D pose is retrieved using RANSAC. This method assumes that the original image and the point cloud are acquired at very close location otherwise the distortion brought by the resizing method would affect the correspondence finding step. Although 2D feature-based methods provided straight forward ways to estimate the optimal alignment between optical image and point cloud, they typically rely on shared information between the two modalities. This can be a major drawback on LiDAR systems are primarily designed to collect range rather than spectral measurements.

3D-based methods offer to align the 3D LiDAR point cloud with the 3D reconstruction of a set of optical images. Corsini et al. propose a two-step method for 3D-based point cloud to image alignment. First, a 3D sparse point cloud is reconstructed from a set of input optical images by using Structure From Motion (SFM) algorithm. The SFM algorithm is designed to find 2D correspondences in images of an input set of images and to regress the 3D pose of each image as well as the 3D position of each feature point, producing a sparse point cloud. After that, the 4-points congruent set (Aiger et al.) algorithm is used to align the sparse 3D point cloud with the 3D LiDAR point cloud. Later, Abayowa et al. propose a similar method for aligning a 3D LiDAR point cloud with a set of aerial optical images. A dense 3D point cloud model is built from the set of optical images using the dense 3D reconstruction method described by Furukawa and Ponce. Then, the pose of the dense point cloud is recovered by using Iterative Closest Point (ICP) (Besl and McKay) algorithm in order to minimize the distance error between the dense point cloud and the LiDAR point cloud. Although these methods achieve high quality results, they require a set of input optical images instead of a single image. Moreover, 3D registration methods are largely sensitive to missing data that often appear in real urban LiDAR data.

Statistical methods for point cloud to image registration try to define metrics that can be used to measure similarities between the two input modalities. Most of the time, the metric is computed in the 2D image domain. The work described in Miled et al. proposes to align the sparse projection of a LiDAR point cloud with an optical image by comparing both modalities using Mutual Information (MI). In Miled et al., this metric is used to find the dependency between the colors carried by the optical images and the reflectances brought by the LiDAR point cloud. The pose between the image and the point cloud is computed using a variational model that maximizes the MI metric between the two modalities. This method achieves very convincing results. However it strongly relies on the quality of the reflectances aquired by the LiDAR sensor. In practical use, only very few high

quality LiDAR sensors can reach such levels of accuracy. Most common sensors acquire reflectance with high level of noise. Moreover, the reflectance is only relevant in certain scenarios and cannot be used on wet surfaces or highly reflective surfaces for example. To overcome the problem of using reflectance, a method for the registration of a raw LiDAR point cloud with a single image is proposed in Castorena et al. (2016). There, the authors propose to fuse and align the modalities at the same time by computing a dense image from the projection of the point cloud and by aligning edges of both modalities. This method can only perform well if the acquisition center of both modalities are very close. Otherwise a lot of ambiguities can arise from the LiDAR projection in the image domain which often leads to large errors in the calibration estimation. Later, (Guislain et al.) proposed a method that aims at aligning only visible points of the LiDAR point cloud with the optical image. To do so, they first estimate the visible points given the optical image point of view using Rubinstein et al. (2008). The remaining points are used to produce a dense image of reflectances by performing bilinear inpainting. This dense reflectance image is aligned with the optical image using a metric that is less sensitive to missing data than Mutual Information. In the case where the reflectance is not available, they offer to compute the same metric on a dense normals map of visible points. This method achieves very good results when the visibility estimation performs well. This is the case when each different objects of the 3D scene are well separated. However, in the case of urban scenes, the amount of missing data as well as the heterogeneity of the shapes and object is very challenging for visibility estimation methods as shown in (Biasutti. et al., 2019). Therefore, the quality of the results on real urban data often lacks of accuracy.

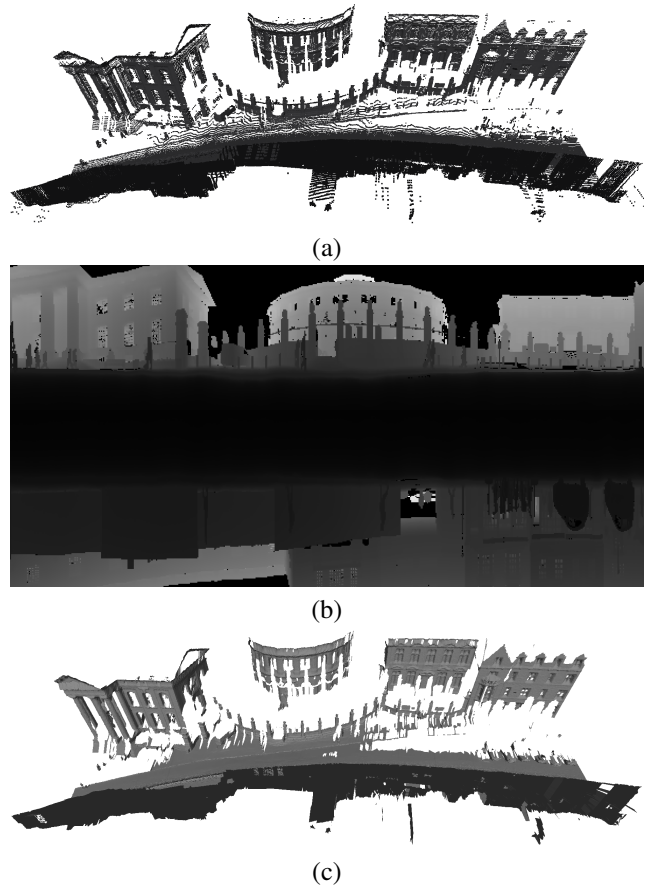

Figure 2: Mesh reconstruction scheme. (a) is the input point cloud, (b) the point cloud as seen in sensor topology and (c) the reconstructed mesh.

In this paper, we only consider the case of fine alignment between a LiDAR point cloud and an optical image. Thus, we estimate that the deformation induced by the perspective between the two modalities can be ignored as the original alignment is close to the optimal one. Therefore, it is sufficient to only estimate a 2D transformation between the optical image and the point cloud in the image domain, which largely simplifies the problem, but limits its usage to close initializations. In the next section, we propose a novel method for LiDAR point cloud and optical image alignment that uses the topology of the LiDAR sensor to generate a dense image without any visibility ambiguities. This dense image is later aligned with the optical image using a variational model.

3. Methodology

In this section, we present each step of the proposed framework for point cloud to image registration. The proposed framework is highlighted Figure 1: first, an image is created by rendering the triangulation based on the sensor topology of the point cloud. Then, this rendering is aligned with the optical image using a variational approach to align the gradients of both modalities.

3.1. Fast mesh reconstruction in sensor topology

The first step of the proposed framework consists in the reconstruction of the mesh of the point cloud. The problem of mesh reconstruction consists in linking points of a point cloud with triangles in order to approximate the surface of the objects in the scene. Surface reconstruction is traditionally done by smoothness approaches (Lipman et al., 2007; Xiong et al., 2014), primitive approximation (Schnabel et al., 2009; Lafarge and Alliez, 2013) or global regularity approaches (Li et al., 2011a,b; Monszpart et al., 2015). However, these methods are often computationally expensive. Moreover, they often require strong assumptions on the homogeneity of the point cloud, which is not suitable in the case of LiDAR acquisitions. To overcome these problems, we propose a very fast approach for mesh reconstruction that exploits sensor topology to instantly create a raw mesh from the point cloud. Note that more precise meshes can be reconstructed using the analogue method proposed in Guinard and Vallet (2018) but with a substantive impact on the computational time. However this work focuses on the efficiency and the performance of the final alignment between LiDAR point cloud and optical image. Thus, the use of the method proposed in Guinard and Vallet (2018) is out of the scope of this work, although it would be interesting to test.

Modern LiDAR sensors often acquire 3D points following a defined pattern from which we can build a dense image (Biasutti et al., 2018). Indeed each point is defined by two angles and a depth, (θ, ϕ, d) respectively, with steps of $(\Delta\theta, \Delta\phi)$ between two consecutive positions. Each point p of the LiDAR point cloud can be mapped to the coordinates (x, y) with $x = \lfloor \frac{\theta}{\Delta\theta} \rfloor, y = \lfloor \frac{\phi}{\Delta\phi} \rfloor$ of a 2D map, hereinafter referred to as u. An example of a point

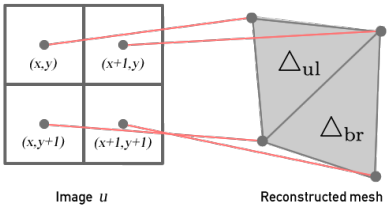

Figure 3: Triangle construction from image in sensor topology.

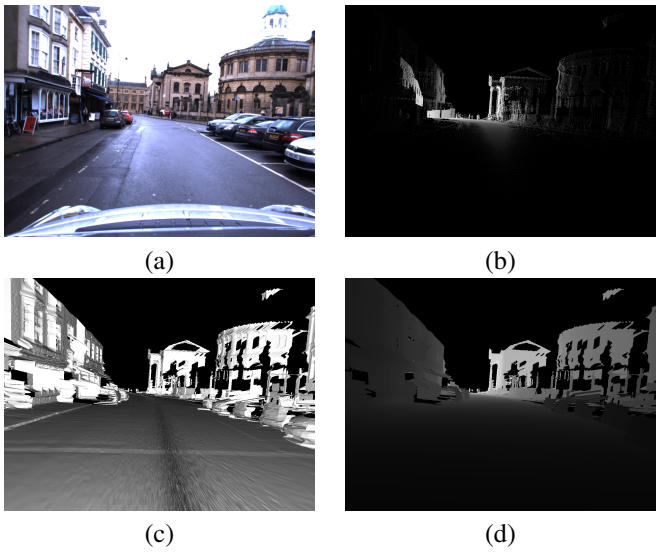

Figure 4: Rendering of mesh at optical image location. (a) is the optical image, (b) is the point cloud projected in the optical image domain without mesh reconstruction, (c) is a texture-less rendering of the mesh and (d) is the depth rendering of the mesh reconstructed from the point cloud.

cloud (Figure 2(a)) as seen from the sensor topology can be met Figure 2(b). This representation of the point cloud enables direct neighborhood computation: the set of neighbors of a given point can be directly retrieved by checking the adjacent pixels of its projection in u.

For each pixel (x, y) of u, 2 triangles $\triangle_{\text{ul}}, \triangle_{\text{bt}}$ are created as follows:

$$\triangle_{\text{ul}} = \{u(x, y),\ u(x+1, y),\ u(x, y+1)\}$$
$$\triangle_{\text{br}} = \{u(x+1, y),\ u(x+1, y+1),\ u(x, y+1)\}$$

This principle is illustrated on Figure 3. After that, triangles are filtered by discarding the ones that have at least one edge that is longer than a certain threshold t, typically $t = 1.0$m. This step prevents separate objects from being connected together which enhances the overall quality of the mesh. An example of reconstructed mesh is showed Figure 2(c). Finally, the mesh is being rendered from the optical camera location, with the same intrinsic parameters. This produces a dense image d_u of the point cloud. As the mesh is not textured, d_u is filled by the values of the *z-buffer* of the rendering (*i.e.* the depth of each pixel). Figure 4 displays an example of a sparse projection of the point cloud (b) in the image domain of (a) compared to texture-less rendering (c) and depth rendering (d). We can see that the renderings are largely denser than the sparse projection, resulting in the appearance of strong depth gradients.

3.2. Depth to optical image alignment

As mentioned in Section 2, the alignment between a LiDAR point cloud $\mathcal{P}$ and an optical image I is non-trivial as both modalities do not share any common attribute. The mesh rendering d_u provides strong depth gradients in the image domain. These gradients correspond to object contours which can also be met in the optical image. Although strong depth gradients can occur without appearing in the optical image, and vice-versa, it is reasonable to assume that most depth gradients also appear in the optical image in real data. Therefore, aligning $\mathcal{P}$ and I in the domain of I can be simplified as the alignment between the gradients of d_u and I. However, this assertion is only true if the initialization of the alignment between d_u and I is relatively close. Indeed, the perspective induced by the 3D rendering introduces deformations that are proportional to the depth of the scene. Thus, if the initialization is too far from the optimal alignment, the alignment between the gradients of d_u and I is not possible.

The method described in Sutour et al. offers to align gradients of two modalities expressed in the same image domain. To that extent, they define a variational model in which gradient alignment between images u_1 and u_2 is done by maximizing the following criterion:

$$C(T) = \int_\Omega \left|\nabla u_1(T_{t_x,t_y,z}(X)) \cdot \nabla u_2(X)\right| dX,$$
$$T_{t_x,t_y,z}(X) = \begin{pmatrix} 1+z & 0 & t_x \\ 0 & 1+z & t_y \\ 0 & 0 & 1 \end{pmatrix} X$$

where Ω is the domain of definition of I and T_{z,t_x,t_y} represents a 2D affine transform with 3 degrees of freedom: vertical and horizontal translation t_x, t_y as well as zooming z. In the case of LiDAR point cloud to optical image alignment, rotation should also be considered in the transform as we cannot assume that the rotation between both sensors is always null. Therefore, we propose to extend the model presented in Sutour et al. in order to estimate rotation as well as translation and zooming.

We define $\bar{T}_{z,t_x,t_y,\theta}$ the 4 degrees of freedom (t_x, t_y translation, z zoom and θ rotation) transformation matrix such that:

$$\bar{T}_{t_x,t_y,z,\theta} = T_{t_x,t_y,z} \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} s\cos\theta & -s\sin\theta & t_x \\ s\sin\theta & s\cos\theta & t_y \\ 0 & 0 & 1 \end{pmatrix}$$

with $s = 1 + z$ to simplify notations. Similarly to Sutour et al., the gradients of d_u and I are aligned by maximizing the following criterion:

$$C(\bar{T}) = \int_\Omega \left|\nabla d_u(\bar{T}_{t_x,t_y,z,\theta}(X)) \cdot \nabla I(X)\right| dX.$$

Using this formulation, an explicit optimization scheme is built to maximize the proposed criterion at each iteration n, by performing a gradient ascent on each parameters of the transformation $\bar{T}_{z,t_x,t_y,\theta}$:

$$\begin{cases} t_x^{n+1} = t_x^n + \lambda_1 \frac{\partial C}{\partial t_x}(\bar{T}_{t_x,t_y,z,\theta}) \\ t_y^{n+1} = t_y^n + \lambda_2 \frac{\partial C}{\partial t_y}(\bar{T}_{t_x,t_y,z,\theta}) \\ z^{n+1} = z^n + \lambda_3 \frac{\partial C}{\partial z}(\bar{T}_{t_x,t_y,z,\theta}) \\ \theta^{n+1} = \theta^n + \lambda_4 \frac{\partial C}{\partial \theta}(\bar{T}_{t_x,t_y,z,\theta}) \end{cases}$$

Table 1: MAE of each method compared to the manually aligned data for each parameter on 50 randomly generated transformations.

Method	Mean Absolute Error			
	t_x	t_y	z	θ
Mutual Information	16.3	11.9	0.05	0.46
Sutour et al. (baseline)	2.91	6.76	0.006	0.57
baseline + rotation	2.96	6.29	**0.004**	0.04
baseline + rotation + refined	**1.93**	**3.31**	0.005	**0.03**

where the partial derivatives of $C(\bar{T}_{t_x,t_y,z,\theta})$ are defined as follows for each iteration:

$$\frac{\partial C}{\partial t_x}(\bar{T}_{t_x,t_y,z,\theta}) = \int_\Omega \sigma \nabla^2 \bar{d}_u(X) \begin{pmatrix} 1 \\ 0 \end{pmatrix} \cdot \nabla I(X) dX,$$
$$\frac{\partial C}{\partial t_y}(\bar{T}_{t_x,t_y,z,\theta}) = \int_\Omega \sigma \nabla^2 \bar{d}_u(X) \begin{pmatrix} 0 \\ 1 \end{pmatrix} \cdot \nabla I(X) dX,$$
$$\frac{\partial C}{\partial z}(\bar{T}_{t_x,t_y,z,\theta}) = \int_\Omega \sigma \nabla^2 \bar{d}_u(X) \begin{pmatrix} x\cos\theta + y\sin\theta \\ -x\sin\theta + y\cos\theta \end{pmatrix} \cdot \nabla I(X) dX,$$
$$\frac{\partial C}{\partial \theta}(\bar{T}_{t_x,t_y,z,\theta}) = \int_\Omega \sigma \nabla^2 \bar{d}_u(X) \begin{pmatrix} -x \cdot s\sin\theta - y \cdot s\cos\theta \\ x \cdot s\cos\theta - y \cdot s\sin\theta \end{pmatrix} \cdot \nabla I(X) dX$$

having $\bar{d}_u(X) = d_u(\bar{T}_{t_x,t_y,z,\theta}(X))$ and $\sigma = \text{sign}(\nabla d_u(X) \cdot \nabla I(X))$. The functional we aim at optimizing is not convex. Therefore, it is highly subject to local maxima. However we consider that the alignment we seek to perform only concerns data provided by calibrated Mobile Mapping Systems. Therefore, the provided alignment of the LiDAR point clouds and the optical images is assumed to be close to the optimal alignment, as discussed here after in Section 4.1.

For the gradient ascent scheme, we set $\lambda_1 = \lambda_2 = 10^{-3}$ to be larger than $\lambda_3 = \lambda_4 = 10^{-5}$ as the translation expressed in pixel is likely to be larger than the rotation or the zooming factor. We set the maximum number of iterations to 200. However, most of our experiments have shown that the method converges in less than 30 iterations on the data presented Section 4.

Finally, we propose to improve the gradient ascent scheme by refining the search steps at each iteration. The search step λ_x^n at iteration n is then defined as follows:

$$\lambda_x^n = \begin{cases} \lambda_x^{n-1} & \text{if} \quad C^n(\bar{T}) > \rho\, C^{n-1}(\bar{T}) \\ \lambda_x^{n-1}/2 & \text{otherwise} \end{cases}$$

with $C^n(\bar{T})$ the energy at iteration n, $\rho = 0.99$. This improvement prevents the algorithm from being directly stuck in a local maxima, and provides better results in practice as demonstrated in Section 4.1.

4. Experiments and results

We conclude this paper by presenting different results obtained using the proposed framework. The proposed pipeline is evaluated on the RobotCar dataset (Maddern et al.) which provides images of resolution 1280×960px as well as point clouds composed of millions of points. We demonstrate the efficiency of the proposed method through a quantitative and qualitative analysis.

4.1. Quantitative analysis

The calibration of the RobotCar dataset does not provide a perfect alignment between LiDAR point clouds and optical images. We propose to manually align mesh renderings with optical images to create ground truths. We found out that the original data alignment compared to the ground truth alignment presents a Mean Absolute Error (MAE) of about 19px for translation, 0.9 degree for rotation and 0.01 for zooming. We propose to apply comparable transformations on the manually align renders to generate evaluation data. The transformations are generated by randomly and uniformly shifting the renders between -20 and 20 pixels on both x and y axis, rotating the renders between -1 and 1 degree and zooming by a uniform random factor between 0.95 and 1.05.

We compare our method with and without the refinement of the search steps to the method proposed in (Sutour et al.), as this method presents the baseline of gradient alignment without the estimation of the rotation. Moreover, we also compare our method to an exhaustive search of the maximum of the Mutual Information (Viola and Wells III) as done in recent multimodal alignment methods, such as (Miled et al.). We compute the MAE between each estimated parameter (t_x, t_y, z, θ) and the ground truth. The results of this experiment are summarized in Table 1. We can see that our method achieves very fine alignment of LiDAR point cloud and optical image. The method with refinement of search steps provides finer results than each other method. The use of the functional defined in (Sutour et al.) as well as the extension presented in this paper outperforms the exhaustive search with Mutual Information metric.

Moreover, we can see that extending the original functional by adding the regression of rotation improves the results not only in the estimation of the rotation, but also in the overall alignment. This is due to the fact that limiting the transformations to translation and scaling prevents the algorithm from finding the optimal alignment. Therefore, the baseline algorithm finds another local maxima which does not align well both modalities. This shows the importance of predicting the rotation as well as the baseline parameters of the transformation. Finally, the refinement of the search steps prevents the variational model from being stuck in local maxima, which makes it more robust to largely shifted initialization while keeping the same computational cost.

4.2. Qualitative analysis

We conclude our experiments with a qualitative analysis. Figure 5 presents the results of LiDAR point cloud to optical image alignment using our method. The first row shows the original alignment, the second row shows the results of the alignments using our method, with closeup looks at the original alignments and our results on the last two rows respectively. On each image, the strong gradients of the depth renderings are represented by green lines on the optical images.

The results presented in Figure 5 highlight that our model succeeds in aligning gradients of both modalities, producing a very good 2D registration between LiDAR point clouds and optical images. From initialization with shifted alignments (shown in the first row), our method produces results where both modalities are seamlessly aligned (second row). In particular, the last row of Figure 5 shows some areas where the variational model perfectly

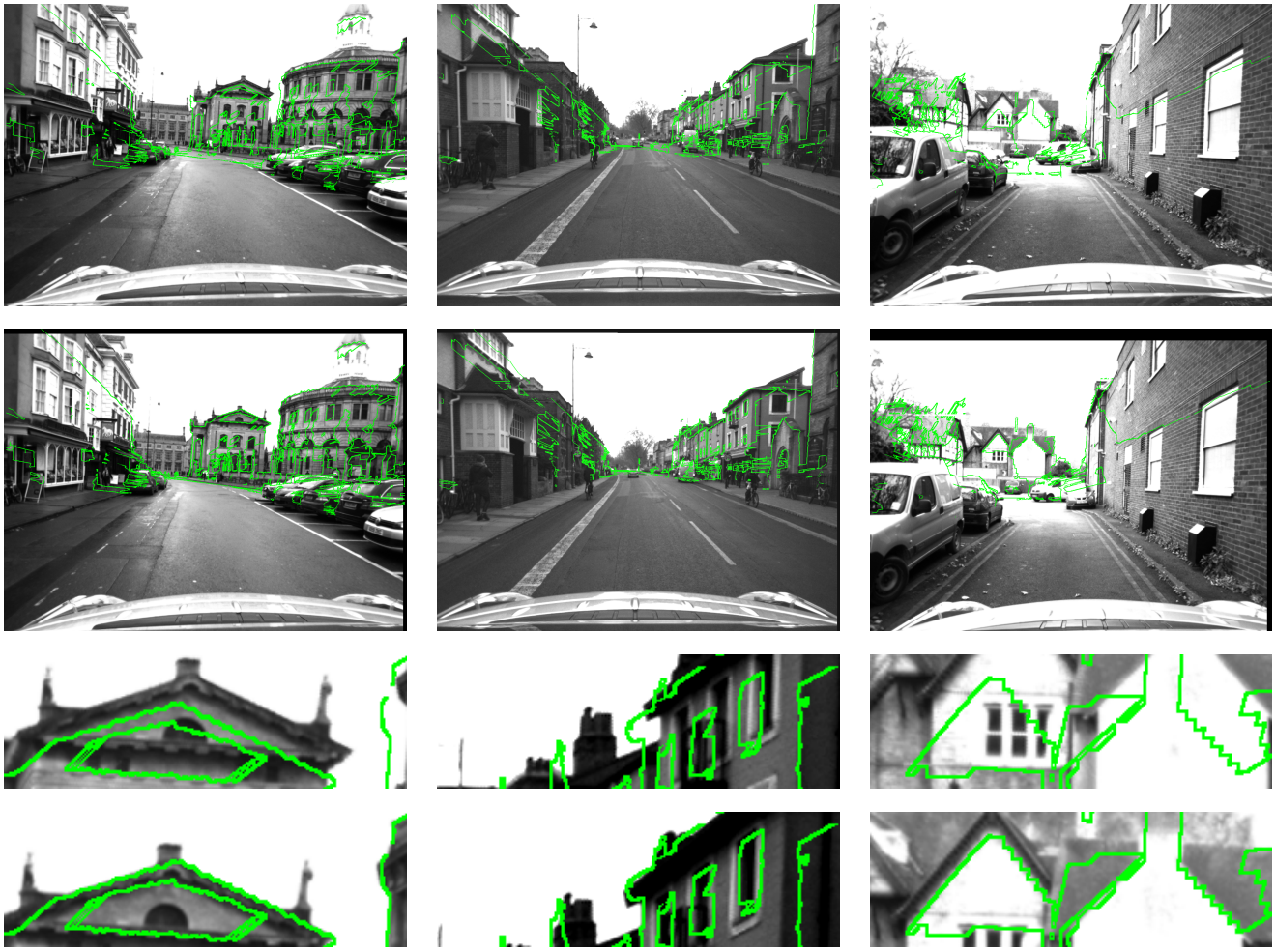

Figure 5: Example of alignments produced by our method. The green lines correspond to the strong gradients of the depth render. The first row shows the original alignment between the optical images and the mesh gradients. The second row shows the alignment produced by our method. A closeup look at details of original alignment and our results is showed on the last two rows respectively.

matches the renders and the optical images on structures that display strong gradients such as roof lines or windows. Moreover, the method only requires to match a small amount of gradients in order to correctly align both modalities. This property makes it more robust to outliers as some gradients of the depth rendering do not correspond to any gradient in the optical image, and vice-versa, as discussed previously in Section 3.2. Finally, our method is able to produce good alignment even when initialized with large shifts between both modalities. This is specially visible in the last column where we can see that in the original alignment, the optical image is shifted from the gradients of the depth render. Despite this initialization, our method succeeds in producing a very fine alignment of the two modalities as it can be seen on the lowest line.

5. Conclusion

In this paper, we have proposed a novel framework for LiDAR point clouds to optical images alignment. The first step of this framework offers to reconstruct the mesh from the point cloud by exploiting the topology of the sensor. After that, the mesh is rendered with the same pose as the optical image. Finally, the depth gradients of the rendered LiDAR mesh and the color gradients of the optical image are aligned using a modified variational approach from Sutour et al.. The qualitative and quantitative results demonstrate that the framework succeeds in very fine alignment between both modalities.

Although the overall results of the proposed method are satisfying, it depends on the initialization of the alignment. In the future, we would like to extend the variational model to perform the gradient alignment from coarse to fine scale to make it less dependant on the initialization. We also would like to compare our current results to the one obtained on renderings done with the mesh produced by the method presented in Guinard and Vallet (2018).

6. Acknowledgement

This project has also received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 777826.

References

Abayowa, B.O., Yilmaz, A., Hardie, R.C., . Automatic registration of optical aerial imagery to a LiDAR point cloud for

generation of city models. ISPRS Journal of Photogrammetry and Remote Sensing 106.

Aiger, D., Mitra, N.J., Cohen-Or, D., . 4-points congruent sets for robust pairwise surface registration, in: ACM Transactions on Graphics.

Allaire, S., Kim, J.J., Breen, S.L., Jaffray, D.A., Pekar, V., 2008. Full orientation invariance and improved feature selectivity of 3D SIFT with application to medical image analysis, in: IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1–8.

Bay, H., Tuytelaars, T., Van Gool, L., 2006. SURF: Speeded Up Robust Features, in: Proc. of ECCV, pp. 404–417.

Besl, P.J., McKay, N.D., . Method for registration of 3D shapes, in: IEEE Trans. on Pattern Analysis and Machine Intelligence.

Biasutti, P., Aujol, J.F., Brédif, M., Bugeau, A., 2018. Range-Image: Incorporating sensor topology for LiDAR point cloud processing. Photogram. Eng. & Remote Sensing 84.

Biasutti., P., Bugeau., A., Aujol., J., Brédif., M., 2019. Visibility Estimation in Point Clouds with Variable Density, in: International Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, pp. 27–35.

Castorena, J., Kamilov, U.S., Boufounos, P.T., 2016. Autocalibration of lidar and optical cameras via edge alignment, in: ICASSP, pp. 2862–2866.

Corsini, M., Dellepiane, M., Ganovelli, F., Gherardi, R., Fusiello, A., Scopigno, R., . Fully automatic registration of image sets on approximate geometry. Int. Jour. of Computer Vision 102.

Fischler, M.A., Bolles, R.C., . Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM 24.

Furukawa, Y., Ponce, J., . Accurate, dense, and robust multiview stereopsis. IEEE Trans. on Pattern Analysis and Machine Intelligence 32.

González, D., Rodríguez-Gonzálvez, P., Gómez-Lahoz, J., . An automatic procedure for co-registration of terrestrial laser scanners and digital cameras. ISPRS Journal of Photogrammetry and Remote Sensing 64.

Guinard, S., Vallet, B., 2018. Sensor-topology based simplicial complex reconstruction from mobile laser scanning IV-2, 121–128.

Guislain, M., Digne, J., Chaine, R., Monnier, G., . Fine scale image registration in large-scale urban LIDAR point sets. Computer Vision and Image Understanding 157.

Lafarge, F., Alliez, P., 2013. Surface reconstruction through point set structuring, in: Computer Graphics Forum, pp. 225–234.

Lepetit, V., Moreno-Noguer, F., Fua, P., . E-PnP: An accurate o(n) solution to the PnP problem. Int. Jour. of Computer Vision 81.

Li, Y., Wu, X., Chrysathou, Y., Sharf, A., Cohen-Or, D., Mitra, N.J., 2011a. Globfit: Consistently fitting primitives by discovering global relations, in: ACM Transactions on Graphics, pp. 52–64.

Li, Y., Zheng, Q., Sharf, A., Cohen-Or, D., Chen, B., Mitra, N.J., 2011b. 2D-3D fusion for layer decomposition of urban facades, in: IEEE Int. Conf. on Computer Vision, pp. 882–889.

Lipman, Y., Cohen-Or, D., Levin, D., Tal-Ezer, H., 2007. Parameterization-free projection for geometry reconstruction.

Lowe, D.G., . Distinctive image features from scale-invariant keypoints. Int. Jour. of Computer Vision 60.

Maddern, W., Pascoe, G., Linegar, C., Newman, P., . 1 Year, 1000km: The Oxford RobotCar Dataset. The International Journal of Robotics Research (IJRR) 36.

Mikolajczyk, K., Schmid, C., . A performance evaluation of local descriptors. IEEE Trans. on Pattern Analysis and Machine Intelligence 27.

Miled, M., Soheilian, B., Habets, E., Vallet, B., . Hybrid online mobile laser scanner calibration through image alignment by mutual information. ISPRS Annals of the Photogrammetry, Remote Sens. and Spatial Inf. Sciences 3.

Monszpart, A., Mellado, N., Brostow, G.J., Mitra, N.J., 2015. RAPter: rebuilding man-made scenes with regular arrangements of planes, in: ACM Transaction on Graphics, pp. 103:1–103:12.

Morel, J.M., Yu, G., . ASIFT: A new framework for fully affine invariant image comparison. SIAM Journal on Imaging Sciences 2.

Moussa, W., Abdel-Wahab, M., Fritsch, D., . An automatic procedure for combining digital images and laser scanner data. ISPRS Annals of the Photogrammetry, Remote Sens. and Spatial Inf. Sciences 39.

Paganelli, C., Peroni, M., Pennati, F., Baroni, G., Summers, P., et al., 2012. Scale Invariant Feature Transform as feature tracking method in 4D imaging: a feasibility study, in: IEEE International Conference of Engineering in Medicine and Biology Society, pp. 6543–6546.

Roshni, V., Revathy, K., 2008. Using mutual information and cross correlation as metrics for registration of images. Journal of Theoretical & Applied Information Technology 4.

Rubinstein, M., Shamir, A., Avidan, S., 2008. Improved seam carving for video retargeting, in: ACM Trans. on Graphics, pp. 16:1–16:9.

Rublee, E., Rabaud, V., Konolige, K., Bradski, G., . Orb: An efficient alternative to sift or surf, in: IEEE Int. Conf. on Computer Vision, pp. 2564–2571.

Schnabel, R., Degener, P., Klein, R., 2009. Completion and reconstruction with primitive shapes, in: Computer Graphics Forum, pp. 503–512.

Sutour, C., Aujol, J.F., Deledalle, C.A., Denis de Senneville, B., . Edge-based multi-modal registration and application for night vision devices. Journal of Mathematical Imaging and Vision 53.

Toews, M., Zöllei, L., Wells, W.M., 2013. Feature-based alignment of volumetric multi-modal images, in: International Conference on Information Processing in Medical Imaging, pp. 25–36.

Viola, P., Wells III, W.M., . Alignment by maximization of mutual information. Int. Jour. of Computer Vision 24.

Xiong, S., Zhang, J., Zheng, J., Cai, J., Liu, L., 2014. Robust surface reconstruction via dictionary learning, in: ACM Transactions on Graphics, pp. 201–205.