Supplementary Material:
Globally Optimal Inlier Set Maximization for Atlanta Frame Estimation

Kyungdon Joo[1] Tae-Hyun Oh[2] In So Kweon[1] Jean-Charles Bazin[3]

[1]RCV Lab., KAIST [2]MIT CSAIL [3]CML, KAIST

{kdjoo369, thoh.mit.edu}@gmail.com {iskweon, bazinjc}@kaist.ac.kr

Summary

This supplementary material contains additional information that could not be included in the main paper due to space limitation:

- Details of efficient bound computation (the relaxed problem and its relationship with the original problem) (Sec. 1)
- Additional experimental results on real-world dataset (Sec. 2)
- Limitations and future work (Sec. 3)

1. Details of Efficient Bound Computation

To speed up the bound computations for large-scale surface normal datasets, we suggest relaxing the inlier region (see Sec. 5.4 in the main paper), inspired by the work of Joo *et al*. [2]. Their method was originally developed in the context of Manhattan frame, but we can adapt it for Atlanta frame using our proposed bound derivations (Sec. 5.2.2 in the main paper) and our Atlanta frame parametrization (Sec. 4.1 in the main paper). We exploit the rectangular bound on the efficient search space (2D EGI)[1] to solve the relaxed problem of Atlanta frame estimation. In this section, we clarify the definition of the relaxed problem and discuss the relationship between the original problem and the relaxed one.

1.1. Relaxed Problem

Let us first recall the original problem formulation (system (3) in the main paper):

$$\underset{\{y\},\mathbf{R}\in SO(3),\{\alpha\}}{\arg\max} \quad \sum_{i=1}^{N}\sum_{j=1}^{M+1} y^{ij} \tag{1a}$$

$$\text{s.t.} \quad y^{ij} d(\mathbf{n}_i, \mathbf{v}_j) \le y^{ij}\tau, \ \forall i, j \tag{1b}$$

$$y^{ij} \in \{0, 1\}, \ \forall i, j. \tag{1c}$$

[1]Extended Gaussian image (EGI) is a kind of surface normal histogram on azimuth and elevation coordinate of the sphere. The EGI on the 3D sphere can be directly transferred to the 2D EGI (*a.k.a.* equirectangular projection).

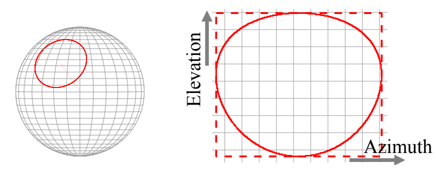

Figure 1: Illustration of bounds (inlier region) on 3D and 2D spaces. *Left*: Inlier region of the original constraint on the sphere. *Right*: Inlier region of the original constraint (solid line) and the relaxed (axis aligned) constraint (dashed line) in the 2D elevation-azimuth space.

We relax the constraints of the original problem as axis-aligned constraints:

$$\underset{\{y\},\mathbf{R}\in SO(3),\{\alpha\}}{\arg\max} \quad \sum_{i=1}^{N}\sum_{j=1}^{M+1} y^{ij} \tag{2a}$$

$$\text{s.t.} \quad y^{ij}\phi(\mathbf{n}_i, \mathbf{v}_j) \le y^{ij}\tau_{\mathrm{el}}, \tag{2b}$$

$$y^{ij}\theta(\mathbf{n}_i, \mathbf{v}_j) \le y^{ij}\tau_{\mathrm{az}}, \tag{2c}$$

$$y^{ij} \in \{0, 1\}, \ \forall i, j, \tag{2d}$$

where $\phi(\cdot,\cdot)$ and $\theta(\cdot,\cdot)$ are the angular distances between two vectors along the elevation and the azimuth axes of 2D EGI, respectively. τ_{el} and τ_{az} are the inlier thresholds for each axis, which are defined by the 3D to 2D EGI mapping. Thanks to the closed-form mapping of 3D to 2D EGI, we can pre-calculate the thresholds τ_{el} and τ_{az}, given the original inlier threshold τ. The visualization of the original and relaxed inlier regions is available in Fig. 1.

With this relaxation, the circular inlier region of the original problem (geodesic distance up to τ on the sphere) is relaxed to a circumscribed rectangular region along azimuth and elevation axes, as shown in Fig. 1. It allows us to leverage the standard integral image techniques on a 2D domain with azimuth and elevation axes rather than 3D sphere. Therefore we can compute, in a constant time, the lower and upper bounds of the number of inliers given a cube of Atlanta directions (systems (8) and (9) in the main paper).

1.2. Relationship with the Original Problem

We now discuss the relationship between the original and relaxed problems (i.e. systems (1) and (2)). Note that this discussion is just for the sake of intellectual curiosity: it is *not* used in the proposed approach, implementation and experiments.

For convenience, we first derive a natural extension of distance function in the underlying metric space. Then, we introduce the Hausdorff distance, which measures the distance between two sets and is used for quantifying solutions obtained from our relaxed problem.

We denote $d(a,b)$ a distance function between two elements a and b. Similarly, we write $d(a,\mathcal{B}) = \inf_{b\in\mathcal{B}} d(a,b)$ a distance between an element a and a set $\mathcal{B}$. Given a distance function d, let $\mathcal{X}$ and $\mathcal{Y}$ be two non-empty subsets of a metric space. The Hausdorff distance, i.e. the distance between two sets, is defined as

$$H_d(\mathcal{X},\mathcal{Y}) = max\left\{\sup_{x\in\mathcal{X}}\inf_{y\in\mathcal{Y}} d(x,y), \sup_{y\in\mathcal{Y}}\inf_{x\in\mathcal{X}} d(x,y)\right\}. \tag{3}$$

The idea of comparing the solutions of the original and relaxed problems is to define distances hierarchically as follows. Using the Hausdorff distance, we measure a distance between two Atlanta frames, and also a distance between two sets of solutions obtained from different problems. An Atlanta frame can be represented as a set of M unit vector directions $\mathcal{V} = \{\mathbf{v}_1,\dots,\mathbf{v}_M\}$. The distance $d_{AF}(\mathcal{V},\mathcal{V}')$ between two Atlanta frames can be computed by Hausdorff distance with geodesic distance as

$$d_{AF}(\mathcal{V},\mathcal{V}') = H_{d_g}(\mathcal{V},\mathcal{V}'), \tag{4}$$

where d_g denotes the geodesic distance between two unit vectors $\mathbf{v}$ and $\mathbf{v}'$ as

$$d_g(\mathbf{v},\mathbf{v}') = \arccos(\mathbf{v}^\top\mathbf{v}'). \tag{5}$$

Given fixed measurements (i.e. input line or surface normals), let $\mathbb{S}$ be the set of solutions obtained by a system. For example, let $\mathbb{S}_o(\tau)$ denotes the set of Atlanta frame solutions obtained by the original system (1) with the inlier threshold τ. Then, the comparison between two solution sets can be made by Hausdorff distance as $H_{d_{AF}}(\mathbb{S},\mathbb{S}')$. Then, we have the following relationship.

Lemma 1. *Given any fixed input measurements, let $\mathbb{S}_o(\tau)$ and $\mathbb{S}_r(\tau)$ respectively be the sets of the globally optimal solutions obtained by solving the original problem (1) and its relaxed version (2), with the inlier threshold τ. For $\epsilon>0$, suppose there exists $\mathbb{S}_o(\tau^\#)$ such that $H_{d_{AF}}(\mathbb{S}_r(\tau),\mathbb{S}_o(\tau^\#)) \le \epsilon$, i.e., ϵ-net, then*

$$\forall V\in\mathbb{S}_r(\tau),\ \left|d_{AF}(V,\mathbb{S}_o(\tau)) - H_{d_{AF}}(\mathbb{S}_o(\tau),\mathbb{S}_o(\tau^\#))\right| \le \epsilon. \tag{6}$$

Proof. We first introduce the properties of the Hausdorff distance $H(\cdot,\cdot)$ [3].

1. If both $\mathcal{X}$ and $\mathcal{Y}$ sets are bounded, then $H(\mathcal{X},\mathcal{Y})$ is guaranteed to be finite.
2. $H(\mathcal{X},\mathcal{Y}) = 0$ iff $\mathcal{X}$ and $\mathcal{Y}$ have the same closure.
3. (Triangle inequality) For every point x of a metric space and any non-empty sets $\mathcal{Y}$, $\mathcal{Z}$ of the same space, $d(x,\mathcal{Y}) \le d(x,\mathcal{Z}) + H(\mathcal{Y},\mathcal{Z})$, where $d(x,\mathcal{Y}) = \inf_{y\in\mathcal{Y}} d(x,y)$

The third property, triangle inequality, holds due to the metric property of the Hausdorff distance. The overall proof is based on this triangle inequality.

We can equivalently write the inequality (6) as follows:

$$\begin{aligned} H_{d_{AF}}(\mathbb{S}_o(\tau),\mathbb{S}_o(\tau^\#)) - \epsilon &\le d_{AF}(V,\mathbb{S}_o(\tau)) \\ &\le H_{d_{AF}}(\mathbb{S}_o(\tau),\mathbb{S}_o(\tau^\#)) + \epsilon. \end{aligned} \tag{7}$$

Let's first prove the right-hand side of (7). Starting from the triangle inequality, we have

$$\begin{aligned} d_{AF}(V,\mathbb{S}_o(\tau)) &\le d_{AF}(V,\mathbb{S}_o(\tau^\#)) + H_{d_{AF}}(\mathbb{S}_o(\tau),\mathbb{S}_o(\tau^\#)) \\ &\le H_{d_{AF}}(\mathbb{S}_r(\tau),\mathbb{S}_o(\tau^\#)) + H_{d_{AF}}(\mathbb{S}_o(\tau),\mathbb{S}_o(\tau^\#)) \\ &\le \epsilon + H_{d_{AF}}(\mathbb{S}_o(\tau),\mathbb{S}_o(\tau^\#)), \end{aligned} \tag{8}$$

where the relationship from the first line to the second line comes from the fact that $\forall x$, $d(x,Y) \le H_{d_{AF}}(X,Y)$, by definition in (3), and the last line is derived by the ϵ-net assumption.

The left-hand side of (7) can be derived from the reverse triangle inequality, which holds for any metric distance.

$$\left|d_{AF}(V,\mathbb{S}_o(\tau^\#)) - H_{d_{AF}}(\mathbb{S}_o(\tau),\mathbb{S}_o(\tau^\#))\right| \le d_{AF}(V,\mathbb{S}_o(\tau)). \tag{9}$$

This can be equivalently expressed as

$$\begin{aligned} -d_{AF}(V,\mathbb{S}_o(\tau)) &\le d_{AF}(V,\mathbb{S}_o(\tau^\#)) - H_{d_{AF}}(\mathbb{S}_o(\tau),\mathbb{S}_o(\tau^\#)) \\ &\le d_{AF}(V,\mathbb{S}_o(\tau)). \end{aligned} \tag{10}$$

We use the inequality between the left-hand side and the middle term.

$$\begin{aligned} -d_{AF}(V,\mathbb{S}_o(\tau)) &\le d_{AF}(V,\mathbb{S}_o(\tau^\#)) - H_{d_{AF}}(\mathbb{S}_o(\tau),\mathbb{S}_o(\tau^\#)) \\ &\le H_{d_{AF}}(\mathbb{S}_r(\tau),\mathbb{S}_o(\tau^\#)) - H_{d_{AF}}(\mathbb{S}_o(\tau),\mathbb{S}_o(\tau^\#)) \\ &\le \epsilon - H_{d_{AF}}(\mathbb{S}_o(\tau),\mathbb{S}_o(\tau^\#)) \end{aligned} \tag{11}$$

By inverting the sign, we have

$$H_{d_{AF}}(\mathbb{S}_o(\tau),\mathbb{S}_o(\tau^\#)) - \epsilon \le d_{AF}(V,\mathbb{S}_o(\tau)). \tag{12}$$

By combining Eqs. (8) and (12), we conclude the proof. □

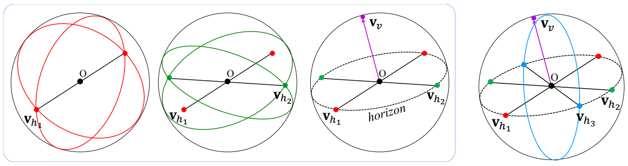

Figure 2: Illustration of the 4-line RANSAC procedure designed for estimating 3 Atlanta directions (left) and example of generalization, 5-line RANSAC for 4 Atlanta directions (right).

Remark Lemma 1 shows that the quality of solutions obtained from the relaxed problem can be quantified by two solutions obtained by the original problem with different inlier thresholds.

2. Experiments

In this section, we will present the details of the generalization of the 4-line RANSAC and show supplementary experimental results on the real-world dataset.

2.1. 4-line RANSAC and its Generalization

In Sec. 6.1 of the main paper, we presented the 4-line RANSAC to detect 3 Atlanta directions. For completeness, we now explain its generalization.

The procedure of 4-line RANSAC is shown in Fig. 2. At each RANSAC iteration, four lines are randomly selected. The idea is to first hypothesize the two horizontal directions, which defines the horizon and thus the vertical direction. The first two lines (shown in red in Fig. 2) intersect at a horizontal direction $\mathbf{v}_{h_1}$ (and its antipodal point). The last two lines (shown in green) intersect at a horizontal direction $\mathbf{v}_{h_2}$ (and its antipodal point). The cross product of $\mathbf{v}_{h_1}$ and $\mathbf{v}_{h_2}$ provides the vertical direction $\mathbf{v}_v$ (shown in purple), which also defines the horizon.

For the generalization of 4-line RANSAC, we can simply define additional horizontal directions $\mathbf{v}_{h_m}$ by computing the intersection point between additional lines and the horizon. For example with 5-line RANSAC, we randomly select 5 lines, and we can find 1 vertical and 3 horizontal directions: we first can generate 3 Atlanta directions $(\mathbf{v}_v, \mathbf{v}_{h_1}, \mathbf{v}_{h_2})$ using the first four lines by 4-line RANSAC, and then define one more horizontal direction $\mathbf{v}_{h_3}$ by computing the intersection between the horizon and the fifth line (in blue), as shown in Fig. 2-right.

2.2. Additional Experimental Results

We now show additional experimental results on real-world datasets that we could not include in the main paper due to space limitations.

York urban database Figs. 4 and 5 show our results with 3 Atlanta directions ($M = 2$) on 60 randomly-selected images from the York urban database. It shows our method can be applied to images with different characteristics such indoor/outdoor urban scenes, low/high number of lines, and different numbers of dominant directions (2 and 3). When the target number of directions (*e.g.* 3) is less than the actual number of directions in the image (*e.g.* 2), then the extra directions are simply clustered to no lines.

Figure 3: Examples of scenes and buildings that do not verify the Atlanta world constraint.

3. Limitations and Future Work

Our approach is designed for Atlanta worlds. Therefore it is not appropriate for scenes and buildings that do not verify the Atlanta world constraint, such as shown in Fig. 3. Similarly, our automatic upright adjustment of VR images requires lines in Atlanta world, so it cannot deal with unstructured landscape pictures, such as the mountain panorama shown in Fig. 3-bottom.

Our approach for line inputs assumes the input images are intrinsically calibrated. An interesting direction for future work would be to estimate the intrinsic parameters of the images, especially the focal length. For example, the goal would be to compute the Atlanta world VPs and the focal length, in such a way that the number of clustered lines is maximized.

References

[1] P. Denis, J. H. Elder, and F. J. Estrada. Efficient edge-based methods for estimating Manhattan frames in urban imagery. In *ECCV*, 2008. 4

[2] K. Joo, T.-H. Oh, J. Kim, and I. S. Kweon. Globally optimal Manhattan frame estimation in real-time. In *CVPR*, 2016. 1

[3] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*. Springer Science & Business Media, 2009. 2

Figure 4: Results of our approach on the York urban database [1]: input lines (odd columns) and our clustering result (even columns).

Figure 5: Same caption as Figure 4.