Interacting Geometric Priors For Robust Multi-Model Fitting: Supplementary Material

Trung T. Pham, Student Member, IEEE, Tat-Jun Chin, Member, IEEE, Konrad Schindler, Senior Member, IEEE and David Suter

I. DETAILED DERIVATION OF EQUATIONS (20), (21)

First, because the function $\delta(.)$ is binary, it is clear that

$$\delta_\beta(t)\delta_\gamma(t) \leq 0.5[\delta_\beta(t) + \delta_\gamma(t)].$$

(1)

And since $C_{\beta,\gamma} \geq 0$ $\forall \beta, \gamma$,

$$C_{\beta,\gamma}\delta_\beta(t)\delta_\gamma(t) \leq 0.5C_{\beta,\gamma}[\delta_\beta(t) + \delta_\gamma(t)].$$

(2)

Summing up all possible pairs $\beta$ and $\gamma$ in inequality (2), we obtain the following inequality

$$\sum_{\beta,\gamma\in\Theta} C_{\beta,\gamma}\delta_\beta(t)\delta_\gamma(t) \leq \sum_{\beta,\gamma\in\Theta} 0.5C_{\beta,\gamma}[\delta_\beta(t) + \delta_\gamma(t)].$$

(3)

As $C_{\beta,\gamma} = C_{\gamma,\beta}$ $\forall \beta, \gamma$, we have:

$$\sum_{\beta,\gamma\in\Theta} C_{\beta,\gamma}[\delta_\beta(t) + \delta_\gamma(t)] = \sum_{\beta,\gamma\in\Theta} [C_{\beta,\gamma}\delta_\beta(t) + C_{\gamma,\beta}\delta_\gamma(t)]$$

$$= \sum_{\beta\in\Theta} \delta_\beta(t) \sum_{\gamma\in\Theta\beta} C_{\beta,\gamma}.$$  

(4)

As a result,

$$\sum_{\beta,\gamma\in\Theta} C_{\beta,\gamma}\delta_\beta(t)\delta_\gamma(t) \leq \sum_{\beta\in\Theta} 0.5\delta_\beta(t) \sum_{\gamma\in\Theta\beta} C_{\beta,\gamma}.$$  

(5)

II. CAN GEOMETRIC PRIORS BE IMPOSED VIA PIECEWISE-SMOOTH MODELS?

On the surface, the geometric consistency could be imposed locally, e.g., neighbouring points are encouraged to be assigned to the consistent models, by using a piecewise-smooth model, e.g.,

$$V(f_p, f_q) = \begin{cases} 
 l + G(f_p, f_q) & \text{if } f_p \neq f_q \\
 0 & \text{otherwise}
\end{cases}.$$  

(6)

where $l$ is a positive spatial smoothness penalty, $G(f_p, f_q)$ is a function measuring the geometric inconsistency between $f_p$ and $f_q$. Under the condition that $l \geq G(\ldots) \geq 0$, function (6) is submodular, and can be minimised by using graph-cut based methods (e.g., PEARL [1]). However, the condition $l \geq G(\ldots) \geq 0$ clearly limits the effect of the geometric consistency since the penalty $G(\ldots)$ is bounded by $l$. Indeed, we have tried PEARL with the piecewise smooth function (6), yet the benefit of the geometric consistency is diminishing. More importantly, in the applications where the smoothness assumption is not valid (e.g., vanishing point detection), it is impossible to penalise the geometric inconsistency using (6) because the smoothness cost is set to zero (i.e., $l = 0$).

III. PLANAR SURFACE RECONSTRUCTION FROM 3D POINT CLOUDS.

Fitting multiple primitives to 3D point cloud has received much interest recently because of the maturity of the 3D reconstruction methods and the increasing popularity of advanced 3D laser scanners. The estimated primitives (e.g., planes, cylinders) can serve as a proxy for 3D scene modelling or meaningful abstractions of the data.

In this section, we aim to fit multiple planes to 3D data which has been reconstructed from a set of images of buildings. We select the publicly available point cloud Merton College 3 from the Oxford Colleges building reconstruction dataset1, and the point clouds used in [2] (specifically Pozzoveggiani and Piazza Dante). These point clouds were obtained using SFM.

Given a set of 3D points $X \subset \mathbb{R}^3$, our goal is to compactly model the data by a small number of planes, where each plane is represented by parameters $\theta = (a, b, c, d) \in \mathbb{R}^4$. The distance (residual) from a point $x_n = (x_n, y_n, z_n)$ to plane $\theta$ is measured as:

$$r(x_n, \theta) = \frac{|ax_n + by_n + cz_n|}{\sqrt{a^2 + b^2 + c^2}}.$$  

(7)

In this application, we exploit the orthogonality and parallelism characteristics of the buildings to enhance the fitting. Accordingly, the geometric consistency between two planes $\beta$ and $\gamma$ is computed as:

$$C(\beta, \gamma) = \begin{cases} 
 0 & \text{if } A(\beta, \gamma) \in [0, 0.5] \cup [89.5, 90] \\
 10 & \text{otherwise},
\end{cases}$$  

(8)

where $A(\beta, \gamma)$ measures the angle between two planes $\beta$ and $\gamma$. A constant per-model cost function is used, i.e., $C(\theta) = 1$, $\forall \theta \in H$.

Parameter settings and hypothesis sampling. The neighbourhood structure $\mathcal{N}$ is obtained by a Delaunay triangulation on $X$. We remove edges that are too long. We set $\sigma = 0.5$, $\lambda = 5$, $K_{max} = 10$. We generate a set $\mathcal{H}$ of 1000 plane hypotheses using the guided sampling method [3]. Note that both Jlinkage and PEARL use the same inlier noise scale $\sigma$ and hypothesis set $\mathcal{H}$ as ours. Other parameters are manually tuned to return the best fitting.

Qualitative comparisons. Fig. 1 provides qualitative comparison results. We choose to plot the top views to easily perceive the angles between the estimated planes. It is clear that the planes fitted by our method more accurately reflect the actual design of the buildings. Observe that although the

1http://www.robots.ox.ac.uk/~vgg/data/data-mview.html
Pozzoveggiani

Piazza Dante

Merton College 3

3D point clouds  Jlinkage  PEARL  MFIGP

Fig. 1. Qualitative comparisons of Jlinkage, PEARL and MFIGP for fitting planes to 3D point clouds. The first column shows the point clouds, and the next three columns display the top views of the fitting results of Jlinkage, PEARL and MFIGP respectively. By observing the angles between the estimated planes, one can appreciate the improvements by our method due to the high-level geometric priors. It can be seen that other methods produce a lot of misalignments of the estimated planes. Remark: arrows point to wrongly aligned planes with respect to the actual building design; detected outliers have been removed from the fitting results for clarity.

data segmentations of Jlinkage and PEARL seem reasonably correct, their fitted planes are considerably biased. This could be explained by the fact that the data errors do not follow a Gaussian distribution. The Merton College 3 dataset (see the bottom left of Fig. 1) clearly demonstrates the case. Under the non-Gaussian noise, the models (planes) with low fitting residuals could be arbitrarily far from the true ones. On the contrary, in our approach, the fitted models balance between the fitting error and the geometric consistency, thus resulting in the better estimations (see the rightmost column of Fig. 1).

IV. CIRCULAR OBJECT DETECTION IN IMAGES

Circular object detection is one of the major challenges in computer vision, in which we are interested in determining the size, location and quantity of objects in the images. The task can be achieved by either fitting circles or disks to the images, depending on the properties of the objects. For example, if the objects of interest have approximately homogeneous intensities, one can fit disks to the image pixels using color models. When the objects have high contrast boundaries with respect to the background and inhomogeneous colors, it is more computationally efficient to fit circles to the extracted edge pixels. We will consider these two models in this experiment.

A circle/disk is parameterized as $\theta = (x_\theta, r_\theta)$ where $x_\theta = (x_\theta, y_\theta)$ and $r_\theta$ are the center and radius, respectively. If $\theta$ is a circle, the distance (residual) between an edge pixel $x_n = (x_n, y_n)$ and $\theta$ is computed as

$$ r(x_n, \theta) = |r_\theta - \sqrt{(x_n - x_\theta)^2 + (y_n - y_\theta)^2}|. \quad (9) $$

If $\theta$ is a disk, the residual is defined as

$$ r(x_n, \theta) = \begin{cases} D(\theta) & \text{if } x_n \in R_\theta \\ \infty & \text{otherwise}, \end{cases} \quad (10) $$

where $R_\theta$ is the region occupied by the object $\theta$ in the image space, $D(\theta)$ measures the quality of $\theta$ using some color models. We use the color model proposed in [4], [5], which is based on the contrast between foreground objects and background.

Here we consider the problem of counting blobs (e.g., cells, flamingoes) in the microscope or overhead satellite images. These objects inherently have circular shapes and are not mutually overlapping (see Fig. 2(a) for an example). Therefore the non-overlapping constraint must be imposed on the object
configurations. In our method, this constraint is enforced via the interactions between any pair objects β and γ, i.e.,
\[
C(\beta, \gamma) = \begin{cases} 
0 & \text{if } \frac{R_{\beta} \cap R_{\gamma}}{\min(R_{\beta}, R_{\gamma})} < 0.1 \\
10 & \text{otherwise.} 
\end{cases}
\] (11)

The per-model cost function is defined as \( C(\theta) = r_{\theta}^2 \) to penalize overly large objects.

A. Fitting circles

Preprocessing. Given an image, we extract a set of edge pixels using Canny detector. These edge pixels are taken as the input data \( \mathcal{X} \). Note that for simplicity we only use the raw edge pixels though the oriented pixels could be used, e.g., for evaluating the spatial smoothness consistency.

Hypothesis sampling. We propose to sample circle candidates \( \mathcal{H} \) using the Hough transform. Specifically, every edge pixel \( x = (x, y) \) votes for all possible bins of a 3D accumulator (the Hough space), in which each bin represents the centre \((x_c, y_c)\) and radius \( r_c \) of a circle. Subsequently we detect 1000 peaks in the Hough space, which are taken as the circle hypotheses. The resolution of 1 pixel for the three axes (i.e., \( x_c, y_c, r_c \)) has been used. We limit the centres to be within the images and radii to be in a range \([20, 100]\) (pruning out too small or too large objects). Note that the peaks are identified...
(a) Input image
(b) Ground truth
(c) MBC
(d) MFIGP
(e) Segmented foreground
(f) PEARL ($w_1 = 0.25$, $w_2 = 0.5$)
(g) PEARL ($w_1 = 0.1$, $w_2 = 0.5$)
(h) PEARL ($w_1 = 0.1$, $w_2 = 1$)

Fig. 4. Counting flamingoes via disk model fitting. (a) is a satellite image of flamingoes [4]. Based on the prior knowledge of the object intensity (i.e., the objects are brighter than the background), we roughly segment the image into the object foreground and background regions (shown in (e)). We then fit disk models to the foreground pixels. Since the image is overhead, the objects should not overlap each others. Thus, MBC and MFIG with the non-overlapping constraint give reasonably better results, as shown in (c) and (d). Although PEARL with the spatial smoothness constraint could penalize the overlapping objects, it is unable to detect small and low-contrasted objects due to the over-smoothing (see (f)). Decreasing the smoothness penalty or increasing the model complexity cost still cannot solve the issue completely (see (g) and (h)).

Ground truth
PEARL
MFIGP
MBC

Fig. 5. Counting cells. From left to right are the ground truth and the detection results of MBC, MFIGP and PEARL. Observe that PEARL tends to merge nearby small cells into a bigger one because of the spatial smoothness constraint. Also overlapping objects cannot be completely avoided by PEARL. On the contrary, MFIGP and MBC result in better detections.

sequentially. Particularly, after the peak with highest value has been detected, the next highest peak not within a radius of 1 from the first peak is found.

**Parameter settings.** We minimise the fitting energy over $\mathcal{H}$ with the following parameters: $\sigma = 1$, $\lambda = 5$, $K_{\text{max}} = 10$. The neighbourhood structure $\mathcal{N}$ is constructed by Delaunay triangulation on $\mathcal{X}$. We remove edges that are longer than 5 pixels.

**Qualitative comparisons.** Fig. 2 depicts the detection results of our approach and other existing multi-model fitting methods such as the Hough transform, AKSWH [6], JLinkage and PEARL using an image of diatom. The result of MBC [4] is also added for comparisons. Note that MBC is a representative of the stochastic geometric framework (see Sec. II in the main manuscript for a review), which fits disk models to the image and imposes the non-overlapping constraint as ours. It can be seen that our method mostly recovers all fully visible objects in the image (see Fig. 2(g)). In contrast, other fitting methods either return redundant objects or miss-detect some true ones. The issues are mainly due to the effects of noise and cluttered outliers. Fig. 2(h) shows the result of MBC, where there are also some false- and miss-detections. This is because the color model used by MBC (i.e., the contrast between the objects and the background) is not sufficiently effective to model the complex background and objects with non-uniform intensities. Fig. 3 demonstrates another result using an image of cells.
B. Fitting disks

Next we test our method for detecting objects which have uniform colors, but are not well contrasted. Also the objects are often highly concentrated, see Fig. 4(a) for a typical image of flamingoes [4]. In such cases, fitting disk models to the image pixels is more appropriate than fitting circles to the edge pixels. However, the downside of this model is that it is computationally expensive. To reduce the computational cost, we roughly pre-segment the image into foreground and background based on the pixel intensity (see Fig. 4(e) for an example). The foreground pixels are then used as the input data $\mathcal{X}$.

**Experimental settings.** To generate the disk hypothesis set $\mathcal{H}$, for each foreground pixel $x$, we sample all possible disks whose centres are at $x$ and radii fall within a range $[r_{\text{min}}, r_{\text{max}}]$. In this application, our method does not enforce the local smoothness constraints since the disk model and the non-overlapping constraints implicitly account for the spatial smoothness. Other parameters are $\sigma = 0.5$, $\lambda = 5$, $K_{\text{max}} = 500$.

We compare our method against PEARL and MBC. For PEARL method, the set of hypotheses $\mathcal{H}$, the residual function, the label cost function and the noise scale $\sigma$ are same as ours. The neighbourhood structure $\mathcal{N}$ is obtained by 4-connectivity. Theoretically, by fitting disks to the image, the strongly overlapping objects are unlikely to be returned since they simultaneously explain the same image region. Also, the spatial smoothness constraint would discourage the overlapping objects since the discontinuity is high when the objects overlap.

**Qualitative comparisons.** Fig. 4 shows the performances of all the tested methods for detecting flamingoes from a satellite image. Our result is clearly comparable with MBC’s result. Figs. 4(f), 4(g) and 4(h) show detection results by PEARL with different values of the weighting parameters $w_1$ and $w_2$ (see Eq. (3) in the main manuscript). As expected, PEARL is able to produce the configuration with non-overlapping objects thanks to the spatial smoothness constraints, however the downside of the spatial smoothness is that some objects are miss-detected due to over-smoothing (the objects are relatively small with respect to the image size) (see Fig. 4(f)). Reducing the smoothness penalty $w_1$ can detect more objects, but redundant objects start appearing (see Fig. 4(g)). Increasing the model complexity weight $w_2$ still can not completely solve the issue (see Fig. 4(h)). Another result using a microscope image of cells [7] is shown in Fig. 5.

**REFERENCES**