ACCV2016
#41

Supplementary Material
Divide and Conquer: Efficient Density-Based Tracking of 3D Sensors in Manhattan World

Yi Zhou [1,2⋆], Laurent Kneip [1,2], Cristian Rodriguez [1,2], Hongdong Li [1,2]

[1] Research School of Engineering, the Australian National University
[2] Australian Centre for Robotic Vision
{yi.zhou, laurent.kneip, cristian.rodriguez, hongdong.li}@anu.edu.au

Abstract. This is the supplemental material for the paper "Divide and Conquer: Efficient Density-Based Tracking of 3D Sensors in Manhattan World". We first introduce the detail about the double Parzen-window based KDE. Then analyze the converging performance of the density distribution alignment. Finally, several simulation experiment and corresponding analysis are provided to demonstrate the robustness and accurateness of our method.

1 Double Parzen-window based KDE

We now explain how to describe the quality of a planar mode in our non-parametric problem by using a double Parzen-window based KDE. We use the 1-D case as an example here.

– Uniform distribution kernel

$$K(x) = \begin{cases} 1 \ |x_i| < \frac{h}{2} \\ 0 \text{ otherwise} \end{cases}$$

– Gaussian kernel

$$K(x) = ce^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1}$$

For a chosen kernel, the corresponding KDE is:

$$p_{KDE}(x) = \frac{1}{Nh^D} \sum_{i=1}^{N} K(\frac{x - x_i}{h}) \tag{2}$$

where D is the dimension of the space which is 1 in this example and 2 in our application, N is the total number of the surface normal vectors which is $640 \times 480 = 307200$ and h is the window size of the kernel.

A statistical measurement λ defined in Eq. 3 is introduced in our non-parametric problem in order to describe the quality of a planar mode. p_u is

⋆ Corresponding author.

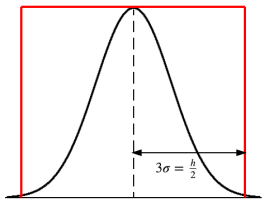


Fig. 1. Double Parzen-window used for calculating the quality of a planar mode. The black one is a Gaussian Kernel while the red one is a uniform kernel.

a KDE with a uniform kernel which represents how likely a surface normal vector locates inside this conic window, namely how dominant a planar mode is. p_g is a KDE with a Gaussian kernel which demonstrates how compact the surface normal vectors surround a mode. We assign c of the Gaussian kernel as 1. Therefore, the more compact that surface normal vectors gather to the mode center the closer that λ approaches 1. The double Parzen-windown based KDE can be understood as a normalized Gaussian KDE which takes into account both the dominance and the accuracy of a planar mode.

$$\lambda = \frac{p_g(x)}{p_u(x)} \tag{3}$$

The complexity of the double parzen-window based KDE is $O(2n)$. Compared to the single window based KDE $O(n)$, it is not a big increase.

2 Convergence Analysis of the Density Distribution Alignment

In order to guarantee the convergence of the minimization of the correlation distance between two discretely sampled distributions f and g, several issues need to be taken into account. First, it is of course vital for f and g to provide density information of the same structure which we call the overlap region. The correlation distance will notably reach its minimum when the overlapping regions align with each other. However, due to the motion of the sensor, the observed structures in successive viewpoints are different, especially along the border of the depth map. This leads to differences in the sampling positions and values. In fact, it is impossible to find the exact overlapping region of a pair of successive depth maps. In our experiment, we find the phenomenon can be greatly weakened by constructing the distribution f and g with only 3D points whose depth is within a proper range (0.5 m $\quad 2 \times median_d - 0.5$ m). Besides, we truncate the distribution f, as shown in Fig 2. This ensures that the sampling positions of the distribution g fully include the ones of the truncated f.

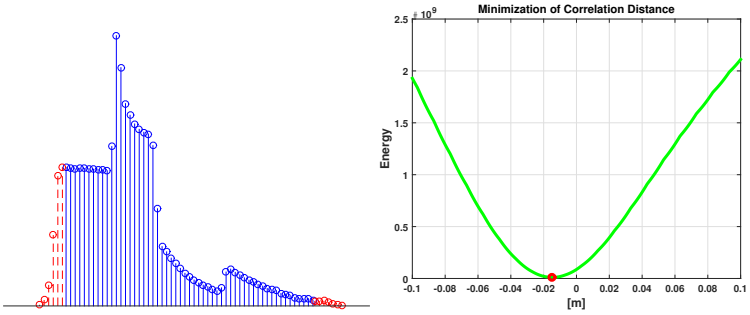


Fig. 2. The left figure shows an example of discretly sampled distribution truncated on the left and right sides (see red dashed lines). The right figure shows the convergence performance. After the truncation, the minimization problem has only one local minimum with a reasonably large convergence basin.

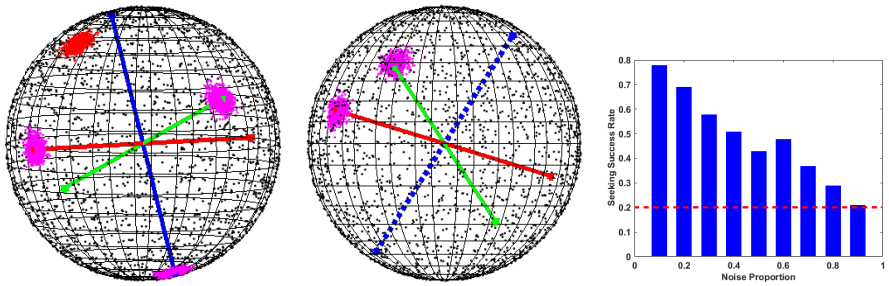
The second issue occurs when the sensor moves orthogonally to the structure, in which case the sampling density changes. A simple solution is to apply a normalization of the distribution. As we observed during experiments on real data, even this is not really needed, except if the sensor moves very close to the structure.

The last issue concerns the choice of the distance function. It is well known that the L_1-distance performs better than the L_2-distance in the presence of outliers. However, there is no noticeable difference in the accuracy of the translation estimation between both norms. This can be attributed to the kernel density distribution alignment, which is robust by nature.

3 Simulation Experiments and Corresponding Analysis

Manhattan frame (MF) seeking in difficult cases In this first simulation experiment we show that our manifold-constrained Manhattan frame tracking (including the initialization) can work robustly in challenging cases that may occur on real data as well:

- In the first experiment, the sensor observes additional planar structures for which the normal vector does not align with any of the Manhattan frame's dominant directions. In this case, there will be more than three modes in the distribution on the unit sphere, as shown in Fig 3 (a). The three cyan modes represent the MF structure while the red one represents an additional slanted plane. Due to the underlying $SO(3)$ manifold-constrained mean-shift updates, which enforce orthogonality in the mode directions, our algorithm ignores the additional mode and converges to the dominant Manhattan frame.
- Another challenging case is that when only two dominant directions of the MF can be observed. In this case, the lost direction can be recovered by exploiting orthogonality and right-handedness between all dominant directions. Fig 3 (b) shows an example of such a situation. Only the two cyan



(a) (b) (c)

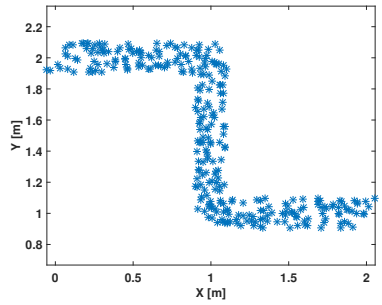
Fig. 3. Robust MF seeking performance in several challenging cases. (a): Seeking the dominant MF when an additional mode/slanted plane exists. (b): Seeking the dominant MF in the case where only two modes can be observed. (c): The success rate of MF seeking under different levels of noise.

modes are found by the algorithm, the third direction (indicated with a blue dotted line) is hallucinated.

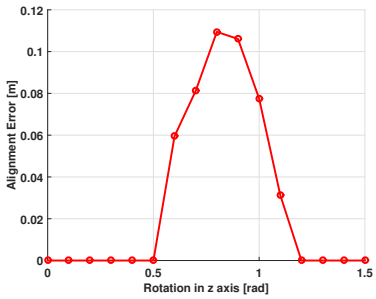
– In Fig 3 (c), we finally demonstrate how the tracking of a MF from a random initial rotation performs under increasing levels of noise. The horizontal axis indicates the overall proportion of noisy normal vectors. It can be observed that as the noise increases, the success rate of the algorithm gradually drops (averaged over many trials). During our initialization procedure, the initial MF orientation is selected from a peak in a histogram over 100 trials. If the overall success rate lies above this peak threshold (0.2 in our experiments), the initial MF is likely to be picked up. Therefore, with 100 trials, our algorithm can successfully initialize the MF even if 90% of the normal vectors represent uniformly distributed noise.

Translation estimation in the Manhattan frame Here we demonstrate the benefit of performing the 1D distribution alignment in the Manhattan frame rather than an arbitrary frame. Without loss of generality, we imagine the two-dimensional example shown in Figure 4 (a). It shows the observation of a simple structure which is perturbed by noise. The structure aligns with the x or y axis of the Manhattan frame. The observation of two arbitrary sensor viewpoints can be simulated by rotating the original structure inside the plane. Figures 4 (c) and (d) show the discrete density distribution along the x-axis of the sensor frame, once from a view-point that is aligned with the Manhattan frame, and once with a rotation of 0.6 rad. It is obvious to see that the distribution inside the Manhattan frame conveys more distinct information than that in an arbitrary sensor view, which is essential for accurate estimation of the translational displacement. The groundtruth displacement in this experiment is 0.1m. Figure 4 (b) illustrates the mean alignment error for different sensor frame orientations (each time averaged over various noise levels). It can be observed that error-free estimation can be

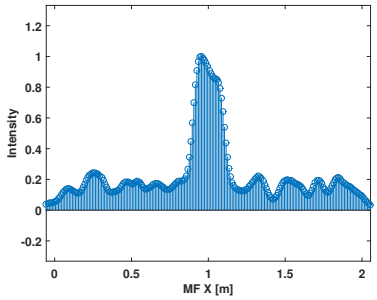
performed if the sensor frame is aligned with the Manhattan frame. In other words, the point cloud needs to be unrotated into the Manhattan frame before establishing the 1D density distribution signals and estimating the translation.



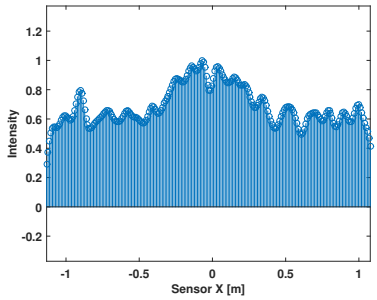
(a) Gaussian noise perturbed structure in MF.



(b) Mean alignment error for different sensor viewing angles.



(c) Density distribution along x-axis of the Manhattan frame.



(d) Denstity distribution along x-axis of a rotated sensor frame.

Fig. 4. Simulation to demonstrate the benefit of performing the distribution alignment in the MF.