Semi-independent Stereo Visual Odometry for Different Field of View Cameras

Trong Phuc Truong[1](✉), Vincent Nozick[1,2](✉), and Hideo Saito[1](✉)

[1] Graduate School of Science and Technology, Keio University, Tokyo, Japan
{ttphuc,saito}@hvrl.ics.keio.ac.jp

[2] Japanese French Laboratory for Informatics, CNRS, UMI 3527, Tokyo, Japan
vincent.nozick@u-pem.fr

Abstract. This paper presents a pipeline for stereo visual odometry using cameras with different fields of view. It gives a proof of concept about how a constraint on the respective field of view of each camera can lead to both an accurate 3D reconstruction and a robust pose estimation. Indeed, when considering a fixed resolution, a narrow field of view has a higher angular resolution and can preserve image texture details. On the other hand, a wide field of view allows to track features over longer periods since the overlap between two successive frames is more substantial. We propose a semi-independent stereo system where each camera performs individually temporal multi-view optimization but their initial parameters are still jointly optimized in an iterative framework. Furthermore, the concept of lead and follow camera is introduced to adaptively propagate information between the cameras. We evaluate the method qualitatively on two indoor datasets, and quantitatively on a synthetic dataset to allow the comparison across different fields of view.

Keywords: Stereo visual odometry · Field of view · 3D reconstruction

1 Introduction

Visual odometry (VO) and simultaneous localization and mapping (SLAM) have been popular research topics in the past decades, and have recently become a prominent part in many emerging technologies such as self-driving car, drone delivery, virtual and augmented reality. Monocular cameras are widely used for these challenging tasks due to their low hardware cost and relatively small size. However, the absolute scale is not observable by using monocular camera approaches without introducing priors, and thus leading to scale drift [4,9]. Stereo camera configurations allow to resolve this scale ambiguity by computing the depth from a known fixed-baseline [12]. In many stereo VO and SLAM using cameras with overlapping fields of view, two identical cameras are often considered to estimate more efficiently correspondences. The second camera is only used to perform static depth estimation [10,13] and/or to add a static constraint

L. Leal-Taixé and S. Roth (Eds.): ECCV 2018 Workshops, LNCS 11129, pp. 430–442, 2019.
https://doi.org/10.1007/978-3-030-11009-3_26

term in the optimization [3,14,17]. Temporal multi-view stereo is thus neglected for the latter since the gain of information would not be worth the computation cost.

This paper presents a proof of concept on how a strong constraint on the focal length difference between the two cameras can result in both a higher reconstruction robustness and accuracy. Indeed, when using cameras with different fields of view, performing temporal multi-view stereo for both cameras can become meaningful as the stereo system will be able to exploit more independent source of data when compared to the case of an identical pair of cameras. In theory, a wider field of view allows to avoid occlusion and it is more likely that the visible part of the scene contains well-suited information for visual methods. Visual odometry and SLAM using large field of view fish-eye cameras [2,15] demonstrate more robust pose estimation, notably during rapid motion, as there is more overlap between subsequent images such that landmarks can be tracked over longer periods. However, the angular resolution of the image decreases as the FOV increases for a fixed image resolution. In [18], Zhang et al. study the impact of the field of view for visual odometry, and show that large field of view camera should be used in confined environment since features are more evenly distributed which stabilizes the pose estimation and can be tracked for a longer time. On the other hand, due to the loss of angular resolution of higher FOV, the triangulation error is amplified with the depth range especially for large scale outdoor environment such that small FOV cameras should be preferred. In this paper, we propose a semi-independent for stereo visual odometry using cameras with different fields of view so that it can take advantage of both the large and small fields of view properties by performing temporal multi-view stereo for both of them.

1.1 Related Work

Using a stereo camera configuration, the scale becomes directly observable given the fixed-baseline, but the implied triangulation can only be estimated for correspondences from both images. As a result, a lot of stereo systems consider a configuration where the common field of view area is maximized.

An early seminal work using a stereo camera setup was proposed by Nister et al. [12], where static triangulation and sequential frame-to-frame matching for sparse features were used to estimate the motion with the correct scale of the stereo rig. In [13], Paz et al. present an approach based on extended Kalman filter that considers information from both close and far features. The former provides scale information through the stereo baseline and the latter are represented with an inverse depth parametrization that is useful to obtain angular information. More recently, Mur-Artal et al. present ORB-SLAM2 [10], an extension of their monocular SLAM framework based on ORB features [9]. The system can work with different configurations such as stereo cameras. It includes loop closing, relocalization, map reuse and follows a similar strategy to [13] by treating differently close and far points.

While these methods are solely based on sparse interest points, recently proposed semi-direct and direct methods have gained popularity due to their ability to circumvent this limitation by exploiting information from all intensity gradients in the image [4,6,11]. Forster et al. present SVO [7], a semi-direct visual odometry, that exploits both photometric error to estimate the initial motion, and geometric error to jointly optimize the camera poses as well as sparse landmarks positions over a window of frames. This method can be easily extended to multiple cameras as the motion estimation and bundle adjustment can be generalized to include measurements from other cameras given their relative pose. On the other hand, full direct methods that only optimize the photometric error also demonstrate state-of-the art results. Based on the work of Engel et al. LSD-SLAM [4], and DSO [3], extension to stereo camera systems have been presented in [5], and [17], in which the authors couple temporal stereo and static stereo in their optimization problem.

1.2 Motivation and Contribution

In this work, we propose a framework for stereo visual odometry using cameras with different yet overlapping fields of view. In particular, we consider the combination of a wide-angle ($\sim$80°) and a medium telephoto lens pinhole camera ($\sim$30°). With this stereo configuration, our system is able to recover the scale by estimating the depth using static stereo matching from the common FOV as illustrated in Fig. 1.

While our method is based on DSO [3], we extend it to work with a stereo configuration such that information between the two different FOV cameras can be shared. Furthermore, it differs from the stereo implementation presented in [17] as we do not directly introduce any constraint from static stereo in the windowed bundle adjustment pipeline. Instead, the back-end optimization is performed individually for each camera as if it were two independent monocular systems to avoid instability that could arise from photometric error depending on the difference of FOV between the cameras. In other words, the temporal multiple-view optimization is performed by both cameras allowing to take advantage of their respective properties; e.g., angular resolution and robust tracking. Furthermore, we introduce an iterative optimization pipeline such that the least reliable camera is initialized in a way that it is more likely to lie in the basin of attraction of the cost function. The front-end part is also modified to initialize the depth variance of each keyframe with static stereo matching and share the depth map used for tracking such that scale drift can be reduced.

Therefore, the proposed method is designated as semi-independent since the two cameras independently execute monocular VO but their initial parameters are jointly optimized. Our main contributions include:

- A stereo visual odometry using different fields of view that can fully exploit, on the one hand, the precision and robustness of the pose of the large FOV camera, and on the other hand, the angular resolution of the small FOV camera, while recovering the reconstruction scale from the known baseline.

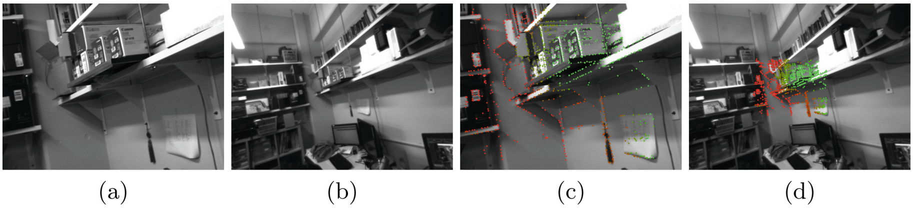
(a) (b) (c) (d)

Fig. 1. Example of stereo image input with different fields of view: (a) 32° and (b) 77°. (c)–(d) Their respective color-coded depth map generated from static stereo matching. (Color figure online)

- An iterative optimization procedure with efficient front-end frame and point management to avoid the joint optimization of two dissimilar cameras.
- Quantitative evaluations on a synthetic dataset with a comparison with DSO and ORB-SLAM2.

2 Stereo Matching

Estimating the depth using images from different physical cameras but taken at the same time (i.e. static stereo) is an important part of stereo VO since it gives information about the scale as the relative position between both cameras is known. Many stereo systems such as [10,14,17] use rectified images as input so that the correspondences search can be performed efficiently along horizontal epipolar lines. However, when considering cameras with notably different focal lengths, there is a loss of the FOV for the wide-lens camera due to distortion or cropping depending on the rectification method used. As a result, rectified images cannot be directly input to our system, but they are still used during static stereo matching for computation time. It can be noted that the 3D point computed from the disparity given rectified images needs to be transformed into the original camera frame since unrectified images are fed as input for the VO pipeline.

Since commonly used matching cost functions were not robust to the difference of resolution between the two rectified images, we define an empirical cost function combining NCC with BRIEF binary descriptor [1]. It allows to avoid local maxima for the NCC by taking into account a sparse but bigger region using BRIEF descriptors. Furthermore, since rectification practically removes any rotation and scale variance, BRIEF provides a good performance under image blur for a low computation time [8]. By defining B_n as the min-max normalized L1 distance between the pixel point p_1 in image I_1 and p_2 in image I_2, and similarly NCC_n as the min-max normalized NCC for the same points, the final cost function C is defined as follows

$$C(p_1, p_2, I_1, I_2) = 1 - B_n(p_1, p_2, I_1, I_2) + \lambda NCC_n(p_1, p_2, I_1, I_2), \qquad (1)$$

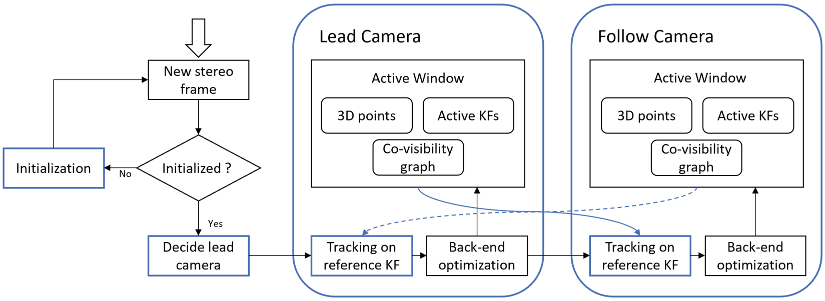

Fig. 2. Overview of our system, blue parts represent our contributions. After initialization, a lead camera is decided and is firstly optimized with the monocular visual odometry pipeline. Given additional information from the active window of the lead camera, the second camera runs, in turn, the monocular pipeline. (Color figure online)

where λ is a weighting factor to balance the influence between NCC and BRIEF. In our experiment, we use NCC with a 5×7 neighborhood, BRIEF with 256 location pairs, and $\lambda = 1$.

3 Stereo VO with Wide and Narrow FOV Cameras

We present a stereo visual odometry method using a wide-angle and a narrow-angle lens camera that combines multi-view stereo from both cameras and static stereo matching from the overlapping FOV. DSO [3] is used as the backbone visual odometry framework since it can benefit from its direct and sparse aspects. In fact, direct method can use every points with high gradient as features so that we can achieve higher resolution point cloud by exploiting the narrow-angle lens camera. Moreover, the sparse nature of DSO allows to save stereo computation time as correspondences are required for a smaller amount of points than dense methods. An overview of our system is presented in Fig. 2, where the blue parts represent our contributions.

3.1 Direct Sparse Odometry Back-End

We adopt the DSO framework as the core visual odometry in our system. DSO proposes a direct probabilistic model with joint optimization of all model parameters including camera poses, camera intrinsic and geometric parameters represented by inverse depths. It is a sparse method that does not incorporate geometric prior so that the Hessian matrix can be solved efficiently using the Schur complement. In [3], the photometric error for a point p in the reference frame I_i, observed in a target frame is defined as the weighted SSD over a 8-point neighborhood $\mathcal{N}_p$ and is formulated as

$$E_{pj} = \sum_{p \in \mathcal{N}_p} w_p \left\| (I_j[p'] - b_j) - \frac{t_j e^{a_j}}{t_i e^{a_i}} (I_i[p] - b_i) \right\|_\gamma , \tag{2}$$

where p' is the warped point of p in I_j; t_i, t_j are the exposure times of the images I_i, I_j; a_i, a_j, b_i, b_j are brightness affine transfer function parameters; $\|.\|_\gamma$ is the Huber norm and w_p is a gradient-dependent weighting defined as

$$w_p = \frac{c^2}{c^2 + \|\nabla I_i(p)\|_2^2}. \tag{3}$$

Each time a new keyframe is created, it is added to the active window which results in an additional energy factor for every points that can be observed by another keyframe in the window as defined (2). The full energy is optimized using Gauss-Newton method, and in order to keep the sliding window of bounded size, marginalization is employed to remove the old keyframes.

3.2 Semi-independent Stereo VO

Iterative Pipeline. Given a stereo configuration with different fields of view, we propose an iterative approach to avoid the uncertainty coming from the difference of resolution during the photometric error optimization. In fact, when comparing the pixel intensity in a reference frame and the one in a target frame from a camera with a different focal length, the impact of noisy pose or depth estimations can result in the non-convergence of the highly complex optimization. The complexity is even more accentuated as we want to perform temporal multi-view stereo for both cameras.

For these reasons, we decouple the problem by performing iteratively two monocular visual odometry pipelines with independent windowed optimizations as illustrated in Fig. 2. At each incoming frame, the most reliable camera (lead camera) is first optimized such that its refined parameters can be thereafter shared with the visual odometry front-end of the other camera (follow camera). As a result, the follow camera, that is considered less reliable, is more likely to converge during its back-end optimization process.

Visual Odometry Front-End. The front-end part of the system handles how frames and points are managed. In particular, it decides which frames and points are added and removed from the windowed optimization. Similarly to DSO, new keyframes are required when the current image becomes too distinctive compared to the last keyframe. It is based on three criteria: when the field of view is significantly different, when the translation part of the motion is high, and when the camera exposure time considerably changes. Each time a keyframe is created, well distributed candidate points with sufficient gradient are selected and their inverse depth variance is directly initialized using static stereo. The front-end also provides initializations for new parameters (camera pose, affine transform parameters, and inverse depth of candidate points) required to optimize the highly non-convex optimization in the windowed optimization.

It differs from stereo DSO [17] by taking advantage of having two semi-independent systems running iteratively. When a new keyframe is created, all active points from both windowed optimizations are projected into the latter and then dilated to create a semi-dense depth map used for tracking the pose of new frames. Since feature points are selected to be well distributed, they can be substantially different for the same area of the two cameras due to the difference of FOV, and thus resulting in denser depth map. This depth map is used to track the camera pose of new frames fed into to the system by minimizing the photometric error using direct image alignment. During this optimization, the inverse depth values are fixed and the two-frame direct image alignment is performed on an image pyramid in a coarse-to-fine order.

Moreover, instead of assuming a constant motion model for the follow system, it is directly initialized using the optimized pose of the lead camera given their constant relative pose. This process is particularly important for the narrow FOV camera as it can easily lose its tracking with respect to the last keyframe during fast motion.

Lead and Follow Camera Selection. The selection of the lead camera is critical as an incorrect pose initialization for the follow camera can result in a divergence from the optimal solution. We propose a straightforward metric to select the lead camera by comparing the latest RMSE results from the windowed optimization. Since the narrow-angle lens camera is more likely to converge to local minimum due to its limited FOV, the latter can become the lead only if the following condition

$$RMSE_{narrow} < f_c RMSE_{wide} \tag{4}$$

is respected 3 keyframes in a row to avoid local minima and maxima results from the optimization. In (4), f_c is a factor to decide which camera should be more trusted. We set $f_c = 0.8$ in our experiment to let the small FOV camera leads the large FOV only when the result of the back-end optimization is 20% lower than the one from the other camera.

3.3 Asynchronous Initialization

Bootstrapping methods for stereo setup based on an initial depth map from static stereo matching as employed in [10,17] will not work efficiently for the wide-angle lens camera. In fact, the estimated depth from static stereo is only limited to the FOV of the narrow-angle lens camera, i.e. the common FOV between the two cameras. We propose an asynchronous method to initialize both cameras in the same coordinate system considering a small overlapping FOV.

Similarly to [17], a semi-dense depth map for the first frame of the small FOV camera can be estimated from static stereo matching to initialize the system.

Once the small FOV system has created N_i keyframes, the corresponding poses for the large FOV system are computed using the relative pose. Then, the point candidates of its first frame can be tracked and their depth values are refined in the subsequent $N_i - 1$ frames by minimizing the photometric error. Moreover, to constraint this discrete search along the epipolar line, all the active points from the small FOV system are projected to the first image plane of the large FOV camera to initialize the associated variance of the candidate points. Finally, the large FOV camera is initialized using the computed depth map and the poses inferred from its counterpart so that both cameras are in the same coordinate system.

4 Evaluation

For the evaluation of our method, we first demonstrate the ability to reconstruct higher resolution point clouds with two indoor datasets, then we evaluate the odometry on a synthetic dataset to be able to compare its accuracy with ORB-SLAM2 and DSO. Since the aim of this paper is to give a proof of concept about stereo systems with different focal lengths, a runtime analysis will not be detailed. However, with an unoptimized implementation, it runs about twice as slow as DSO considering it has to compute a second time the back-end optimization and estimate stereo matches each time a keyframe is generated.

4.1 Point Cloud Reconstruction

The stereo configuration used to evaluate the point cloud accuracy is a 77° and 32° FOV camera with a ~20 cm baseline. The two datasets contain 800 frames representing around 15 s of video of an indoor environment. Some examples of input images are illustrated in Fig. 3. It also shows the estimated trajectory of the camera pair as well as the point cloud generated using our semi-independent visual odometry method. The accuracy of the registration of both system can be observed by comparing the 3D points from the large and small FOV camera represented in red and green respectively. The color mapping of the trajectory shows that for these two datasets, the camera lead was the narrow-angle lens one. It can be explained by the fact that the motion was relatively slow. However, for both datasets, the lead switches to the big FOV because of the sudden change of direction. The last row illustrates the difference of density between the 3D reconstruction of each camera. In particular, it shows that using a medium telephoto camera, the point selection of DSO can focus more on specific details of the scene. The point cloud is thus more detailed even if the camera pair is at a reasonable distance from the scene.

4.2 Odometry Accuracy

We evaluate our method on the Urban Canyon model [18], where photorealistic synthetic images were generated for a stereo pinhole camera setup with different

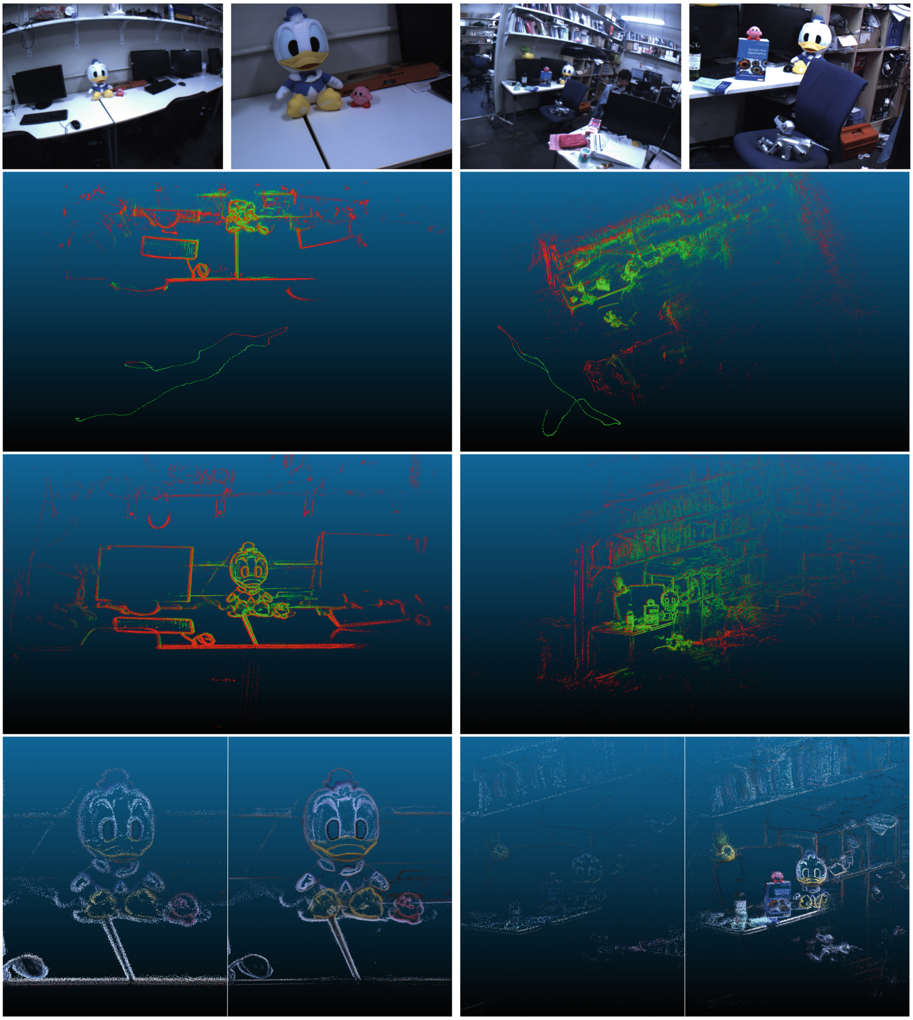

Fig. 3. Qualitative examples on two indoor datasets. (*First row*) Example of input images. (*Second and third row*) 3D reconstruction of the large and small FOV camera are shown in *red* and *green* respectively. The camera trajectory is also represented on the *top row*, and the same color mapping represents which camera was the lead. (*Last row*) A zoomed view of the 3D reconstruction of the large FOV camera displayed on the *left* and the small one on the *right* (Color figure online)

FOV (40°, 60°, and 80°) by using cycle raytracing engine implemented in Blender (Fig. 4). This way, many stereo configurations for the same trajectory can be proposed to study the impact of the field of view. We compare the accuracy of the visual odometry with ORB-SLAM2 and monocular DSO (since the stereo version is not available publicly). The trajectories are aligned to the ground truth using a rigid-body transform (6DoF) for the stereo methods and a similarity transform (7DoF) for DSO. It can be noted that to allow a fair comparison between all methods, loop closure and relocalization were disabled for ORB-SLAM2, we also disable real-time forcing and we use the default parameters for DSO and ORB-SLAM2. We use three different metrics proposed in [16] for our evaluation: the absolute translation RMSE t_{abs}, the relative translation RMSE t_{rel}, and the average relative rotation error r_{rel}.

In the remaining part of this paper, we denote the different results as follows:

- *Ours40*: our semi-independent stereo VO using 40°-80° FOV stereo camera
- *Ours60*: our semi-independent stereo VO using 60°-80° FOV stereo camera
- *DSO60*: DSO using 60° monocular camera
- *DSO80*: DSO using 80° monocular camera
- *ORB40*: ORB-SLAM2 using 40° stereo camera
- *ORB80*: ORB-SLAM2 using 80° stereo camera

Table 1 summarizes the visual odometry results for the different evaluated configurations. Their trajectory can be observed on Fig. 5. While it does not prove the versatility of our method, *Ours40* and *Ours60* have the best results for the absolute trajectory error and relative rotation error. The reason is that it is able to exploit the wide FOV camera and it slightly outperforms *DSO80* for the absolute trajectory error since information from the narrow FOV is also exploited. While ORB-SLAM2 manages to estimate correctly the relative rotation, the translational error is higher than the two other methods. A reason could be that ORB features are not suitable for this synthetic data since increasing the number of feature points resulted in higher errors in our experiment.

Table 1. Comparison of accuracy in the Urban Canyon dataset. t_{abs} absolute translation RMSE (m), t_{rel} relative translation RMSE (%), r_{rel} average relative rotation error (deg/10 m). Best results are shown as bold numbers

	t_{abs}	t_{rel}	r_{rel}
Ours40	**0.275**	1.236	0.896
Ours60	0.428	3.918	**0.486**
DSO60	0.906	1.352	0.612
DSO80	0.292	**0.439**	0.622
ORB40	1.599	2.120	0.533
ORB80	0.929	1.856	0.543

5 Discussion and Conclusion

In most of stereo VO and SLAM methods, homogeneous camera are considered to take advantage of their overlapping fields of view and the ability to efficiently

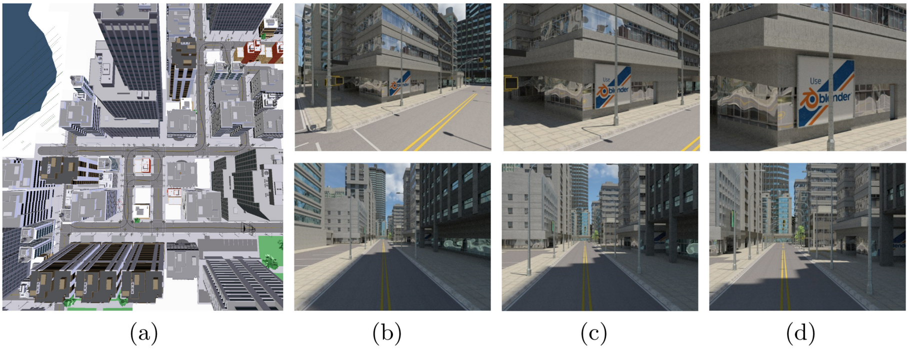
(a) (b) (c) (d)

Fig. 4. (a) Top view of the Urban Canyon 3D model. Examples of synthetic images with different fields of view: (b) 80° (c) 60° (d) 40°, each row corresponds to the same camera position.

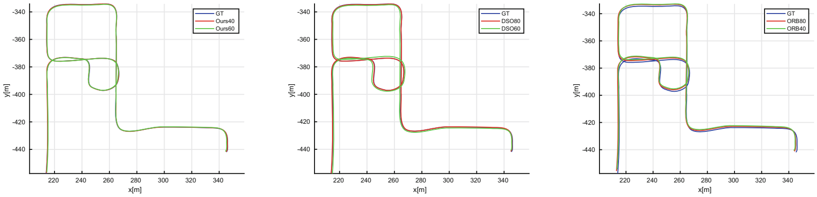

Fig. 5. Qualitative results on the Urban Canyon dataset. (*From left to right*) Trajectory of our semi-independent method, monocular DSO, and stereo ORB-SLAM2

estimate matches. As a result, temporal information from the second camera is often omitted since it does not provide additional meaningful data to the stereo system when compared to the first camera. In this paper, we suggest that, by using heterogeneous stereo camera with different focal lengths, performing temporal multi-view stereo optimization for both cameras can lead to better 3D reconstruction while having a robust pose estimation. This proof of concept is illustrated by our semi-independent stereo visual odometry for large FOV and small FOV cameras. Some preliminary results show the ability to reconstruct high detailed 3D point clouds while standing at a reasonable distance and to estimate with accuracy the camera pose when compared to DSO and ORB-SLAM2 for the proposed synthetic dataset.

While it does not prove that our method is constantly better, it exposes the ability to choose different focal lengths for a multiple cameras setup. For example, this stereo configuration is already present in many smartphones to allow depth of field rendering. Nevertheless, because of the limited range of the static stereo depth estimation due to the small baseline and the difference of FOV, most of common stereo VO and SLAM methods are not suitable. In this case, employing a semi-independent approach could result in a better performance.

References

1. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: BRIEF: binary robust independent elementary features. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6314, pp. 778–792. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15561-1_56
2. Caruso, D., Engel, J., Cremers, D.: Large-scale direct SLAM for omnidirectional cameras. In: IROS, vol. 1, p. 2 (2015)
3. Engel, J., Koltun, V., Cremers, D.: Direct sparse odometry. IEEE Trans. Pattern Anal. Mach. Intell. **4** (2017)
4. Engel, J., Schöps, T., Cremers, D.: LSD-SLAM: large-scale direct monocular SLAM. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8690, pp. 834–849. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10605-2_54
5. Engel, J., Stückler, J., Cremers, D.: Large-scale direct slam with stereo cameras. In: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1935–1942. IEEE (2015)
6. Forster, C., Pizzoli, M., Scaramuzza, D.: SVO: fast semi-direct monocular visual odometry. In: 2014 IEEE International Conference on Robotics and Automation (ICRA), pp. 15–22. IEEE (2014)
7. Forster, C., Zhang, Z., Gassner, M., Werlberger, M., Scaramuzza, D.: SVO: semidirect visual odometry for monocular and multicamera systems. IEEE Trans. Rob. **33**(2), 249–265 (2017)
8. Heinly, J., Dunn, E., Frahm, J.-M.: Comparative evaluation of binary features. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, pp. 759–773. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33709-3_54
9. Mur-Artal, R., Montiel, J.M.M., Tardos, J.D.: ORB-SLAM: a versatile and accurate monocular SLAM system. IEEE Trans. Rob. **31**(5), 1147–1163 (2015)
10. Mur-Artal, R., Tardós, J.D.: ORB-SLAM2: an open-source slam system for monocular, stereo, and RGB-D cameras. IEEE Trans. Rob. **33**(5), 1255–1262 (2017)
11. Newcombe, R.A., Lovegrove, S.J., Davison, A.J.: DTAM: dense tracking and mapping in real-time. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 2320–2327. IEEE (2011)
12. Nistér, D., Naroditsky, O., Bergen, J.: Visual odometry. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004, vol. 1, p. I. IEEE (2004)
13. Paz, L.M., Piniés, P., Tardós, J.D., Neira, J.: Large-scale 6-DOF SLAM with stereo-in-hand. IEEE Trans. Rob. **24**(5), 946–957 (2008)
14. Pire, T., Fischer, T., Civera, J., De Cristóforis, P., Berlles, J.J.: Stereo parallel tracking and mapping for robot localization. In: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1373–1378. IEEE (2015)
15. Rituerto, A., Puig, L., Guerrero, J.: Comparison of omnidirectional and conventional monocular systems for visual SLAM. In: 10th OMNIVIS with RSS (2010)
16. Sturm, J., Engelhard, N., Endres, F., Burgard, W., Cremers, D.: A benchmark for the evaluation of RGB-D SLAM systems. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 573–580. IEEE (2012)

17. Wang, R., Schwörer, M., Cremers, D.: Stereo DSO: large-scale direct sparse visual odometry with stereo cameras. In: International Conference on Computer Vision (ICCV), vol. 42 (2017)
18. Zhang, Z., Rebecq, H., Forster, C., Scaramuzza, D.: Benefit of large field-of-view cameras for visual odometry. In: 2016 IEEE International Conference on Robotics and Automation (ICRA), pp. 801–808. IEEE (2016)