Tackling the Scale Factor Issue in a Monocular Visual Odometry Using a 3D City Model

Paul Verlaine Gakne and Kyle O'Keefe,

Position, Location And Navigation (PLAN) Group
Department of Geomatics Engineering
Schulich School of Engineering
University of Calgary
Alberta, Canada
{pvgakne, kpgokeef}@ucalgary.ca

BIOGRAPHY

Paul Verlaine Gakne is a PhD candidate in the Position, Location And Navigation (PLAN) Group in the Department of Geomatics Engineering at the University of Calgary. He received his MSc in Communication and Information Systems from Huazhong University of Science and Technology, Wuhan, PR China and BSc in Telecommunications from Ecole Supérieure Multinationale des Télécommunications, Dakar, Senegal (Cameroon branch). His research focuses on sensors fusion and satellite-based navigation for vehicular applications.

Kyle O’Keefe is a Professor of Geomatics Engineering at the University of Calgary, in Calgary, Alberta, Canada. He has worked in positioning and navigation research since 1996 and in satellite navigation since 1998. His major research interests are GNSS system simulation and assessment, space applications of GNSS, carrier phase positioning, and local, indoor, and vehicular navigation with ground based ranging systems and other sensors.

1. INTRODUCTION

In urban canyons, Global Navigation Satellite System (GNSS) users are surrounded by tall buildings. In such areas, there are three sub-groups of GNSS signals: blocked signals, shadowed signals, and direct signals. As a result, GNSS receivers are subject to multipath errors (line-of-sight (LOS) signal received along with non-line-of-sight (NLOS) signal). In this scenario, the GNSS receiver will perform poorly often with several tens of meters of position error.

In order to mitigate the multipath error and increase urban canyon user accuracy, various research approaches have been explored over past years. Existing methods focus mainly on integrating GNSS with other sensors (Inertial Measurement Units (IMU), Camera, 3D building model, etc.)

This paper focuses on using an existing 3D building model (city of Calgary, Alberta, Canada) to compute a scale factor that will be used to scale the translation computed from a monocular visual odometry system. This is done by using the slant distance obtained from the skyline matching between the camera and 3D building model synthesized images.

2. OBJECTIVES

Monocular-based positioning for pedestrian [1] or vehicular applications [2] has been previously studied by exploiting various concepts such as the rigid-body motion concept [3], and the use of feature points [4]. With a stereo-camera system, it is possible to directly obtain the platform's true translation magnitude by triangulating feature points and obtaining the depth information. However, this is still a challenge for a monocular system. Significant research has been done to define the road as a plane and use the vehicle height to obtain the scale information [5, 6]. However, these approaches still contain errors because the assumption that the road is a plane is only true up to an extent. In this paper, we will use a special setup with a sky-pointing camera whose captured images will be used in two ways: (*i*). for visual odometry; (*ii*). for skyline-based positioning.

The objective of this paper is twofold:

1. To integrate the position information obtained from a 3D building model by matching the camera images with synthesized 3D building model images with visual odometry.
2. To calculate the scale factor from the 3D building model to increase the translation magnitude accuracy

The 3D building model-based positioning used in point 1 of the objective has been realized in our previous work [7]. In this work, the output of this algorithm will be tightly integrated with the visual odometry output.

3. METHODOLOGY

The methodology used in this paper will follow the following steps:

3.1 Monocular visual odometry

The monocular visual odometry developed here is based on feature point detection, matching then finally estimating the vehicle motion.

An example of feature detection, matching and outlier rejection is given in Figure 1.

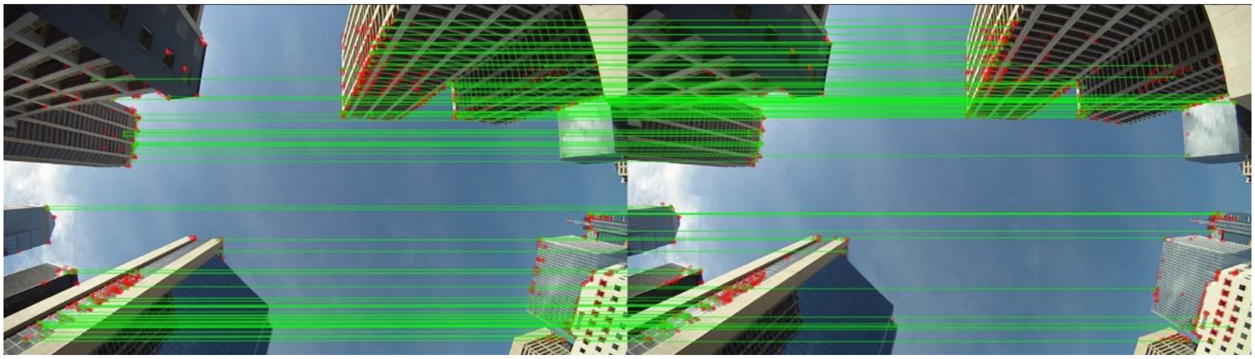


Figure 1: Feature point detection, description (ORB algorithm), matching and outlier removal (RANSAC). Left: image frame at time t; Right: image frame at time t + Δt. Red circles represent feature points that are detected but not used (outlier); Green circles are those that are detected and properly matched between consecutive image frames – established matches illustrated by the green lines

Given two sets of M feature points $f_{p_1} = \{p_0, p_1, \dots, p_{M-1}\}$ and $f_{p_2} = \{p'_0, p'_1, \dots, p'_{M-1}\}$, the rotation and translation can be determined using least-squares and approach using the singular value decomposition (SVD). The rotation and translation (T_c) of the camera (R_c) is calculated such that:

$$(R_c, T_c) = \arg\min_{r_c, t_c} \sum_{j=0}^{M-1} w_j ||(r_c p_j + t_c) - p'_j||^2 \quad (1)$$

Where $w_j > 0$ is the weight of each point pair.

The steps summarizing the feature-based visual odometry are depicted in Figure 2.

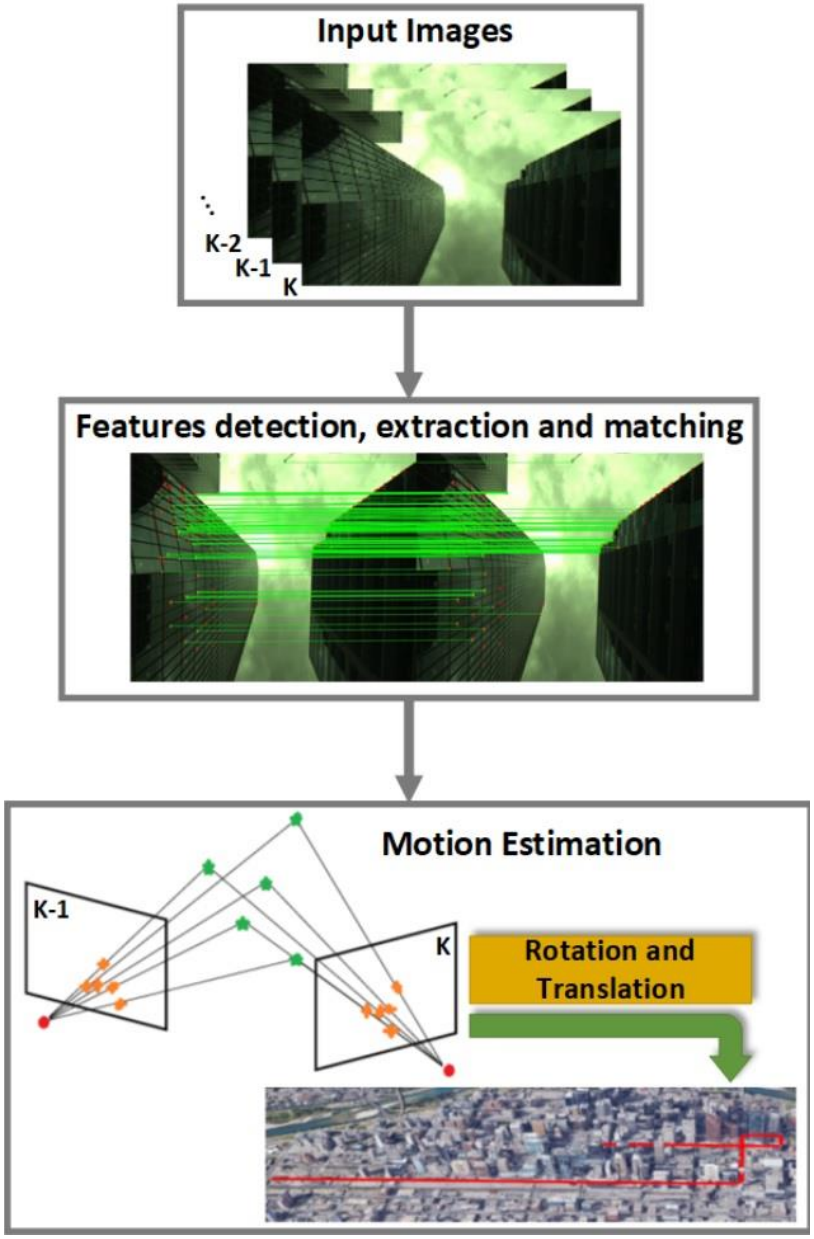


Figure 2: Feature points-based visual odometry

3.2 3D building-based Positioning

The second important concept used in this paper is 3D building positioning using images captured from the monocular system. The localization problem can be solved using skylines obtained from a 3D city model and from an upward-facing camera. A visual spectrum narrow field-of-view camera similar to that available on most mobile phones is used. The corrected images are then segmented into sky and non-sky areas. The obtained binary images are compared with the ideal images synthetized from the 3D building model. The position solution estimate corresponds to the best match obtained between the observed and synthetized images. The process of obtaining the vehicle location based on the 3D building model is given in Figure 3.

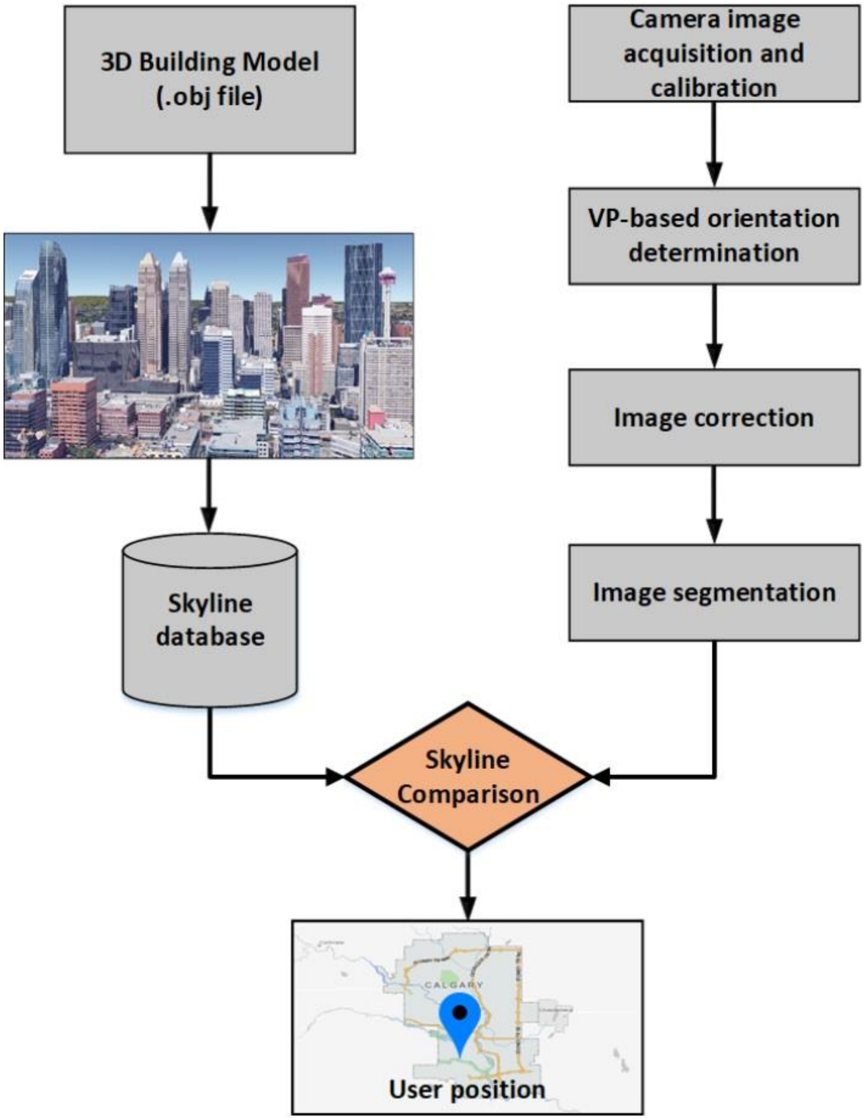
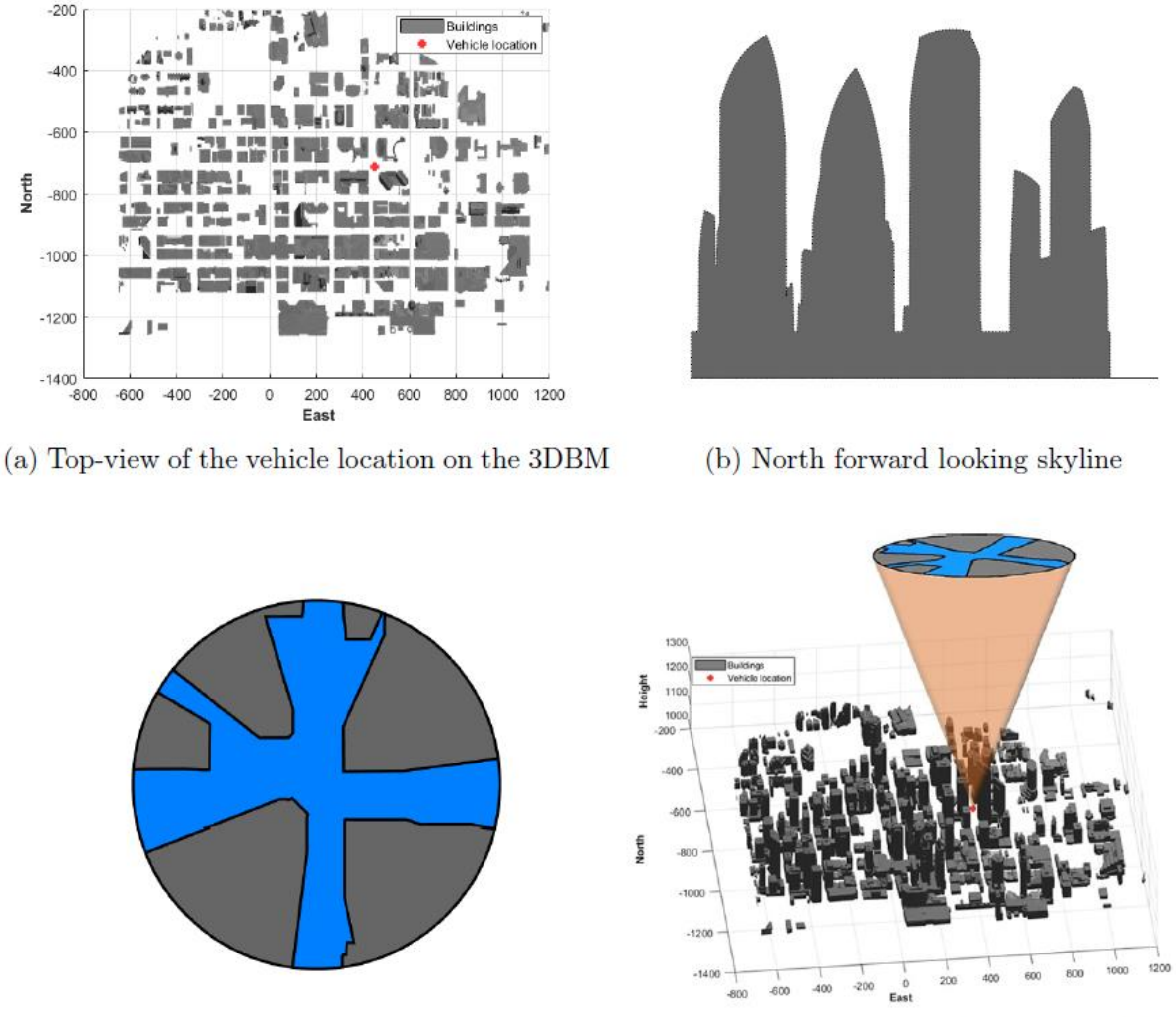

3D Building Model (.obj file)
Skyline database
Camera image acquisition and calibration
VP-based orientation determination
Image correction
Image segmentation
Skyline Comparison
CALGARY
User position

Figure 3: Skyline-based positioning flowchart

At each vehicle location, the skyline is computed from the 3D building model over a range of azimuth (Figure 4). The skyline is defined as:

$$\mathbb{C}_{\mathbf{p}} = \{(\boldsymbol{\alpha}_{\mathbf{p,i}}, \mathbf{h}_{\mathbf{p,i}})\} \text{ for } \mathbf{i} = \mathbf{0..N-1} \tag{2}$$

where $\mathbf{i}$ and $\mathbf{N}$ are the azimuth index and number respectively. $\boldsymbol{\alpha}$ is the highest elevation angle of the obstructing surface and $\mathbf{h}$ its corresponding height. An azimuth resolution of 0.5° is used in this work i.e., $\mathbf{N} = \mathbf{720}$ points to define the skyline.



(a) Top-view of the vehicle location on the 3DBM

(b) North forward looking skyline

(c) Upward looking skyline

(d) Location of the vehicle on the 3DBM

Figure 4: Location of the vehicle indicated on the 3DBM as well as the ideal synthesized forward and upward skylines at the same location.

Example of the segmented camera image and the synthesized 3D building images are depicted in Figure 5 and 6 respectively. Details on the step-by-step determination of each binary images will be provided in full version of the paper.

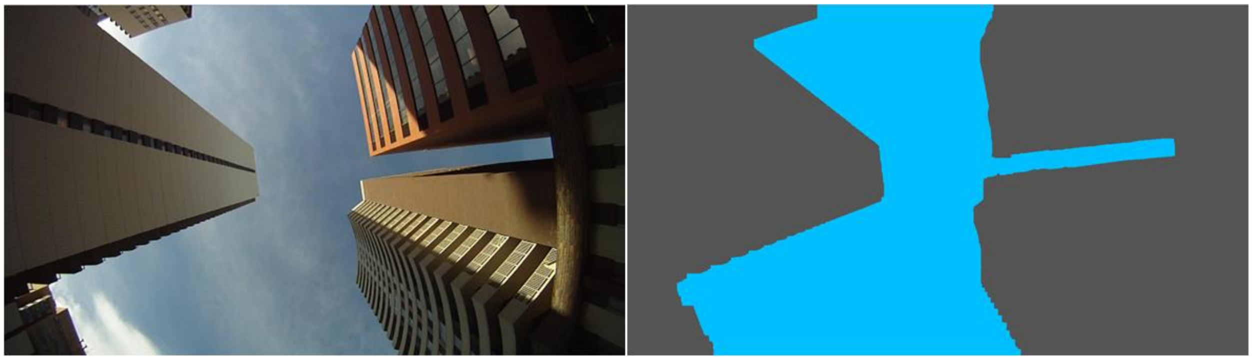


Figure 5: Example of the camera image. Left: original image, right: segmented image

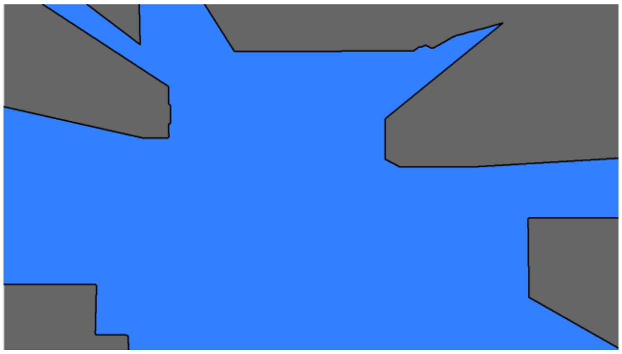


Figure 6: Example of a 3D building model synthesized image

The binary images are then compared using a similarity metric. Herein we use the cross-correlation coefficient (CC) defined as:

$$\mathbf{CC}_{\mathbf{I}_{b(p)}}(\boldsymbol{x},\boldsymbol{y}) = \frac{\mathbf{1}}{\mathbf{n_{px}}}\sum\nolimits_{x,y}([\mathbf{I}_{\boldsymbol{b}}]_{\mathbf{obs}}(\boldsymbol{x},\boldsymbol{y}) \circ [\mathbf{I}_{\boldsymbol{b}}]_{\mathbf{3D}_{(p)}}(\boldsymbol{x},\boldsymbol{y})) \tag{3}$$

where p is the database image's position vector; ° represents the Hadamard product of two matrices; $\mathbf{CC}_{\mathbf{I}_b}$ is the cross-correlation coefficient of the binary images; $[\mathbf{I}_{\boldsymbol{b}}]_{\mathbf{obs}}$ is the observed binarized image; $[\mathbf{I}_{\boldsymbol{b}}]_{\mathbf{3D}}$ represents the synthetized binary image from the 3D building model. The similarity metric obtained as in Figure 7 gives the location of the vehicle on the travelled path.

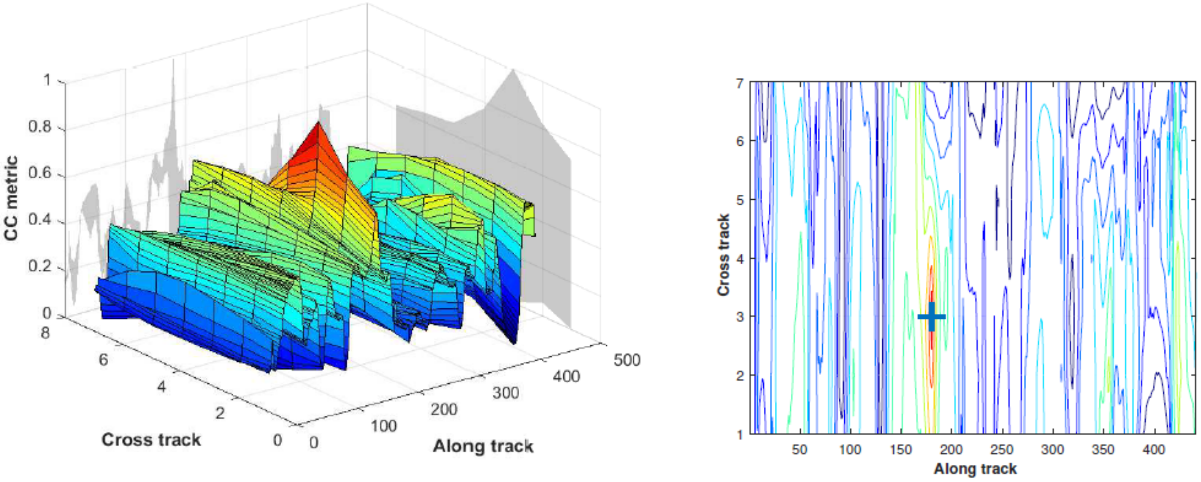


Figure 7: similarity metric indicating the location of the vehicle on the road

By this process, the position error is computed as depicted in Figure 8.

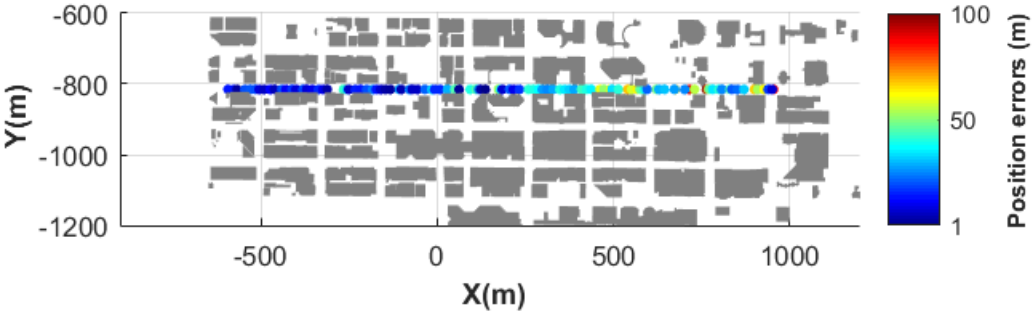


Figure 8: Position error along one Avenue

From the skyline defined as in Equation (2), the slant distance to a point (the highest point in view from the vehicle location is chosen) is computed as:

$$d_s = \frac{h}{\sin\alpha} \tag{4}$$

Where h and α are defined as in Equation (2).

Then the scale factor is computed as:

$$s = \frac{d_s}{||P_{3DBM} - P_{VO}||} \quad (5)$$

Where P_{VO} represents the position obtained from the visual odometry and P_{3DBM} the position obtained from the 3D building model.

This information is then used to scale the translation obtained from the visual odometry calculated as in Equation (1).

4. RESEARCH CONTRIBUTIONS

This paper will present the follow novel contributions:

- **Camera rotation correction using Vanishing Points (VPs)**: the concept of vanishing points, wherein parallel lines in the real world appear to converge to a point in images can be used to determine the image inclination and rectify the captured images in order to increase the skyline matching accuracy.
- **Image segmentation**: synthetized images (from the 3D building model) are ideal images since they do not suffer from any lighting conditions or occlusion by nearer to the camera objects. However, images from upward-pointing cameras taken in day-time can be challenging to segment. This means that the skyline obtained from the camera can be severely degraded and inaccurate. As a result, comparing skylines (database versus camera images) may not be helpful for positioning. Thus, the paper will present an improved image segmentation method for sky and obstacle (non-sky, e.g., building) identification.
- **Scale factor computation**: a scale factor is computed from a monocular system to emulate the information obtained from sensors such as LiDAR to aid the visual odometry and improve the final vehicle location computation.

5. ON-GOING WORK AND INITIAL RESULTS

At this point, data are collected, and initial tests are complete. Initial results are embedded in the methodology section. More details regarding each step and the integration of the 3D building model solution with the visual odometry will be provided in the final version of the paper.

6. REFERENCES

[1]. Marouane C., Gutschale R., Linnhoff-Popien C. (2018), “Visual Odometry for Pedestrians Based on Orientation Attributes of SURF.” In: Bi Y., Kapoor S., Bhatia R. (eds) Proceedings of SAI Intelligent Systems Conference (IntelliSys) 2016. IntelliSys 2016. Lecture Notes in Networks and Systems, vol 16. Springer, Cham

[2]. Aumayer, B. M. (2016). “Ultra-Tightly Coupled Vision/GNSS for Automotive Applications.” Department of Geomatics Engineering, University of Calgary, Calgary.

[3]. Ma, Y., S. Soatto, J. Kosecka, and S. S. Sastry (2003). “An Invitation to 3-D Vision: From Images to Geometric Models.” SpringerVerlag.

[4]. Lowe, D. G. (2004). “Distinctive image features from scale-invariant keypoints.” International Journal of Computer Vision 60, 91-110.

[5]. S. Choi, J. Park and W. Yu, "Resolving scale ambiguity for monocular visual odometry," 2013 10th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), Jeju, 2013, pp. 604-608.

[6]. Kitt, B.; Rehder, J.; Chambers, A.; Schönbein, M.; Lategahn, H.; Singh, S. (2011), "Monocular Visual Odometry using a Planar Road Model to Solve Scale Ambiguity." Proceedings of the 5th European Conference on Mobile Robots (ECMR 2011), Örebro, Sweden, September 7-9, 2011. Ed.: A. J. Lilienthal

[7]. Gakne, Paul Verlaine, O'Keefe, Kyle, "Skyline-based Positioning in Urban Canyons Using a Narrow FOV Upward-Facing Camera," Proceedings of the 30th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2017), Portland, Oregon, September 2017, pp. 2574-2586.