Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnLine 49-15 (2016) 108–113

Ground plane based visual odometry for RGBD-Cameras using orthogonal projection

Julian Jordan * **Andreas Zell** *

* *Department of Computer Science, Faculty of Science, University of Tuebingen, Tuebingen, 72076 Germany (e-mail: {julian.jordan, anderas.zell}@uni-tuebingen.de).*

Abstract: This work presents a method for visual odometry that allows robust 3 degrees of freedom trajectory estimation for wheeled robots using a downward facing RGBD-camera. Assuming that the robot moves on a ground plane while the environment itself can have arbitrary geometry allows to estimate the frame to frame motion from orthographic projections of the RGBD-data. Instead of directly aligning these projections, the reference frame is split into blocks, which are individually registered using Efficient Second Order Minimization, and thus create several estimates of the current motion. These estimates are combined using an outlier rejection scheme to create a robust estimate of the actual motion even under challenging conditions. The results of this method are compared to the results of other state-of-the-art methods to show its accuracy and robustness.

Keywords: Mobile robots, Visual, Motion estimation, Image registration

1. INTRODUCTION

For electric wheelchairs the usage of RGBD-sensors is quite common for obstacle detection and localization, for example Kırcalı and Tek (2014), Wei et al. (2013) and Wu et al. (2013). We want to provide comparable sensor capabilities to users of an electric wheeled walker. Since these need to be more lightweight and also are more restricted in terms of maximum cost, this should be achieved using a single RGBD-sensor. As this sensor is also used for obstacle detection it has to be downward facing in order to see obstacles that are below the ground plane, such as stairs. Therefore, most information visible to the RGBD-camera is the floor, which in many cases like in homes for the elderly or hospitals, has few objects on it and a repetitive low contrast texture. Varying lighting conditions and image acquisition problems like over/under exposure, motion blur and image noise create a very challenging environment for visual odometry. But there is one advantage in this scenario: due to the nature of the platform and its users we can assume that they move in a nearly planar environment with almost no slope, allowing us to reduce the number of degrees-of-freedom to three. In combination with the availability of depth information, this allows the creation of an orthographic projection of the environment. That is sufficient for estimating the frame-to-frame motion in 3DoF while being more robust than a full 6DoF approach. The proposed method consists of four processing steps: First, an orthographic projection of the current RGBD data is created. Then the projection of the previous frame is split into a number of image blocks. These blocks are registered using Efficient Second order Minimization (ESM), each of them giving an estimate of the global robot motion. The final motion estimate is calculated by removing outliers from the block estimates and combining them using a weighting function. All these steps can be performed in real-time on a midrange CPU.

2. RELATED WORK

Many different approaches have been presented that solve the odometry estimation problem using cameras. There are two main classes of visual odometry systems: methods that use features extraction at key points to find point correspondences (feature based methods) and methods that are estimating the motion by minimizing the photometric error between two images (direct methods). The planar assumption, describing a vehicle that performs planar motion parallel to a ground plane, is often used to estimate the ego-motion of a car. Adding these constraints reduces the number of parameters which need to be estimated and therefore simplifies the calculation. Additionally, rectifying the image with regard to the ground plane allows to directly estimate the motion parameters by using image alignment, either by aligning the complete image or several sub patches. The approach presented in (Mano and Shashua, 2000) uses alignment of rectified image patches and selecting a subset, based on geometric and photometric constraints, of these patches to increase robustness. Using a virtual downward looking camera and its advantages is described in (Ke and Kanade, 2003), besides that this approach also utilizes patch registration and selection. A hybrid method using feature correspondences and direct alignment is described in (Azuma et al., 2010). (Lovegrove et al., 2011) shows that whole image direct VO (visual odometry) for an on-road vehicle is possible using a downward facing camera capturing the planar road surface, but also describes due to the lack of outlier rejection, two major problems of the proposed method. In

2405-8963 © 2016, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2016.07.717

(Zienkiewicz and Davison, 2014) this approach is extended and shown that it also works for indoor robots on various materials with different texture and reflection properties. A method using registration of image patches created from a fronto-parallel projection is described in (Kitt et al., 2011), but due to the way these patches are registered only small rotations are allowed. Extracting features on the ground plane is used in (Caglioti and Gasparini, 2007) to estimate robot motion. Feature extraction is also used in (Scaramuzza et al., 2009), but by applying car specific motion constraints a single feature correspondence is sufficient to perform visual odometry. In (Hamme et al., 2015) a similar method is presented that uses back projection to track features on the ground plane instead in image space. A general direct visual odometry method using RGB-D data is presented in (Kerl et al., 2013b) and later extended to a full SLAM system: (Kerl et al., 2013a). In (Klose et al., 2013) three different image alignment methods are evaluated: Forward Compositional (FC), Inverse Compositional (IC) and Efficient Second order Minimization (ESM). They also show that adding a global affine illumination term to the optimisation improves the performance of all three methods. These three methods are based on the iterative image alignment approach presented in (Lucas and Kanade, 1981). The image alignment method used in our work uses ESM as described in (Benhimane and Malis, 2004), extended with a global illumination term.

3. PROPOSED METHOD

Our method describes the visual odometry problem as an iterative process that estimates the camera motion by finding the image warp that transforms one frame to the next.

Our motion model assumes that the robot is moving parallel to the ground plane ($z = 0$ in world coordinates) and therefore the frame to frame motion can be described by three parameters: $x \in \mathbb{R}^3 = (\Delta x, \Delta y, \Delta\theta)^T$. Further we assume that the camera is mounted on the robot in a known fixed pose C. ${}^{C}T_L$ transforms this pose to the robots local camera coordinate system L, which has a X-Y plane equal the X-Y plane of the world coordinate system.

${}^{C}T_L$ is defined by three parameters roll ϕ_c, pitch ψ_c and height h_c. In our experiments these parameters were calibrated by placing the robot on a flat surface and fitting a plane to the point cloud produced by the RGBD-sensor. To correctly model the robot's motion we also need to know its center B and the transform ${}^{L}T_B$, which also has to be calibrated or measured externally. The robot's location is described by the transform ${}^{W}T_B$ and is the product of the inter frame motion ${}^{B'}T_B$. At a certain time t, with a current frame I_c and robot pose B and a previous frame I_r at time t' and robot Pose B', ${}^{W}T_B$ is defined as:

$$ {}^{W}T_B = {}^{W}T_{B'} {}^{B'}T_B \tag{1} $$

3.1 Orthographic Projection

As an input for our algorithm we use the point cloud PC acquired by the RGBD sensor. Since the used image registration method only operates on intensities, the RGB information is reduced to intensity values. An example

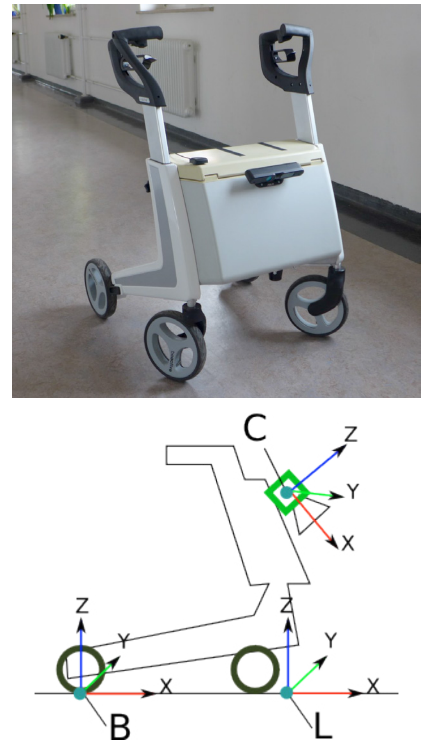

Fig. 1. Top: Our prototype system. Bottom: Schematic drawing with frames B, C and L .

orthographic projection image is shown in Fig.2 (top image).

To create the orthogonal projection we first transform PC to the robots local camera coordinate system L:

$$ PC' = ({}^{C}T_L)PC \tag{2} $$

Let I_z be the orthogonal projection image and $p = ([x, y, z]^T, i) \in PC'$ a point, where i is the point's intensity. The intensity of a pixel p' in I_z is:

$$ p' = \frac{\sum_p^{PC'} (w(p, p')p.i)}{\sum_p^{PC'} w(p, p')} \tag{3} $$

with

$$ w(p, p') = \begin{cases} f(p, p') & \text{if } |p.x - p'.x| \leq 1.0 \text{ and} \\ & |p.y - p'.y| \leq 1.0 \\ 0 & \text{otherwise.} \end{cases} $$

and

$$ f(p, p') = (1.0 - |p.x - p'.x|)(1.0 - |p.y - p'.y|) \tag{4} $$

This results in a 2D image containing the intensities of all points present in the current area of interest. Since some pixels in I_z do not get any intensity values assigned we need to exclude them from further processing. This is done using a mask image I_m, containing 0 if a pixel's weight $w(p, p')$ is below a certain threshold t_a and 1 otherwise.

4. IMAGE REGISTRATION

For camera based odometry to work, the photo consistency assumption is required (Kerl et al., 2013a): Two images of the same point p taken with a camera at two different positions m_1 and m_2 have the same intensity: $I(x) = I'(x')$ with x, x' being the pixel positions of p in image I and I'.