**Formula derivation of Direct Sparse Visual-Inertial Odometry with Stereo Cameras**

Ziqiang Wang

March, 2018

Ziqiang Wang is an engineer of UISEE Technologies Shanghai Co., Ltd and was a postgraduate student with the college of Electronics and Information Engineering of Tongji Universty, Shanghai China.(e-mail: ziqiang.wang@ uisee.com).

**ABSTRACT**

We present Stereo VI-DSO, a novel approach for visualinertial odometry, which jointly estimates camera poses and sparse scene geometry by minimizing photometric and IMU measurement errors in a combined energy functional. The visual part of the system performs a bundle-adjustment like optimization on a sparse set of points, but unlike key-point based systems it directly minimizes a photometric error. This makes it possible for the system to track not only corners, but any pixels with large enough intensity gradients. IMU information is accumulated between several frames using measurement preintegration, and is inserted into the optimization as an additional constraint between keyframes. We explicitly include scale and gravity direction into our model and jointly optimize them together with other variables such as poses. As the scale is often not immediately observable using IMU data this allows us to initialize our visual-inertial system with an arbitrary scale instead of having to delay the initialization until everything is observable. We perform partial marginalization of old variables so that updates can be computed in a reasonable time. In order to keep the system consistent we propose a novel strategy which we call ”dynamic marginalization”. This technique allows us to use partial marginalization even in cases where the initial scale estimate is far from the optimum. We evaluate our method on the challenging EuRoC dataset, showing that VI-DSO outperforms the state of the art.

**Key Words:** VSLAM, IMU, Stereo vision, Mobile robot

Content

[Chapter1 PHOTO RESIDUALS 2](#_Toc531252386)

[1.1 INTRODUCTION 2](#_Toc531252387)

[1.1.1 NOTATION 2](#_Toc531252388)

[1.1.2 QUESTION IMPORT 2](#_Toc531252389)

[1.2 SOLUTION 3](#_Toc531252390)

[1.2.1 CONSTRUCT RESIDUAL 3](#_Toc531252391)

[1.2.2 JACOBIAN CITATION 5](#_Toc531252392)

[1.2.3 JACOBIAN DERIVATION 5](#_Toc531252393)

# Chapter1 PHOTO RESIDUALS

## 1.1 INTRODUCTION

Windowed Optimization is a classic method in non-linear optimization.

### 1.1.1 NOTATION

Throughout the paper, we will write matrices as bold capital letters () and vectors as bold lower case letters (), light lower-case letters to denote scalars (). Light upper-case letters are used to represent functions ().

Homogeneous camera calibration matrices are denoted by  as (2.1). Camera poses are represented by matrices of the special Euclidean group , which transform a 3D coordinate from the camera coordinate system to the world coordinate system. In this paper, a homogeneous 2D image coordinate point  is represented by its image coordinate and inverse depth as (2.1) relative to its host keyframe . The host keyframe is the frame the point got selected from. Corresponding homogeneous 3D world coordinate point  is denoted as (2.1).  are used to denote camera projection functions. The jacobian of ,  is denoted as (2.1)

### 1.1.2 QUESTION IMPORT

Assume we observe 5 points  in 4 keyframes , every keyframe has stereo vision  abbreviated as . A point can also be observed by other frame as shown in Table(2.1). Question is how to use Windowed Optimization method to make our observation more accurate ?

Table (2.1)

|  |  |  |
| --- | --- | --- |
| Image point | Host keyframe | Observe by |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

## 1.2 SOLUTION

We use direct method to construct residual, Windowed Gauss-Newton method to optimization residual。

### 1.2.1 CONSTRUCT RESIDUAL

Dynamic multi-view stereo residuals  are defined as

 is Huber norm.  is affine brightness parameters to frame  .  is a gradient-dependent weighting parameters,  in frame  projected to  is  as:

Static one-view stereo residuals  are modified to

Hostframe of  is .  is affine brightness parameters to frame .  in frame  projected to  is  as :

Total residuals

To balance the relative weights of temporal multi-view and static stereo, we introduce a coupling factor  to weight the constraints from static stereo differently.  is a set of all image point host by frame .  are the observations of  from temporal multi-view stereo. If there are  image point and  keyframes in , optimization variable  is

In this example, there are 7 dynamic residuals and 3 static residuals, Factor graph of the residuals function is

Total residuals is



Iteration  can be calculated by

We construct residuals and its formulation.

### 1.2.2 JACOBIAN CITATION

 We know for a Lie algebra  and :

### 1.2.3 JACOBIAN DERIVATION

#### 1.2.3.1 Dynamic Parameter

Firstly, if  is neither observed by frame ,  nor hosted by , :

otherwise, we follow

For one frame , we have  and , then we can get

Secondly, according to

We have:

add detail Calibration derivation……



#### 1.2.3.2 Static Parameter

Firstly, For a stereo frame : inverse depth , a left frame  pixel  is projected to right frame  with :



Secondly, according to:

We have:

add detail Calibration derivation……