**Supplementary Material to:**

**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).

Content

[Chapter1 PHOTO RESIDUALS 3](#_Toc531252386)

[1.1 INTRODUCTION 3](#_Toc531252387)

[1.1.1 NOTATION 3](#_Toc531252388)

[1.1.2 QUESTION IMPORT 3](#_Toc531252389)

[1.2 SOLUTION 4](#_Toc531252390)

[1.2.1 CONSTRUCT RESIDUAL 4](#_Toc531252391)

[1.2.2 JACOBIAN CITATION 6](#_Toc531252392)

[1.2.3 JACOBIAN DERIVATION 6](#_Toc531252393)

# Chapter1 VISUAL-INERTIAL PRELIMINARIES

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 (), typewriter letters are used to represent functions ().

Rigid-body orientation directly is described as elements of  and poses as . We can identify every skew symmetric matrix with a vector in  using the hat  operator [2, eq. (1)]. Exponential map associates Lie Algebra to a pose and logarithm is anti-mapping following [2, eq. (3)]:

.

The term , is the right Jacobian of ,.We write directly as vectors, i.e.,  and . we use the right perturbation retraction for ,



and for , we perturb transformation on the right,



The input for our Stereo VI-DSO is a stream of IMU measurements and stereo camera frames. In IMU body fame(abbreviated as “”), the gyroscope and accelerometer measurements at time , namely  and  , are affected by additive white noise  and a slowly varying sensor bias . is sampling intervals. The state of IMU at time  is described by the orientation, position, velocity from “” to the world frame “” and biases:



Velocities live in a vector space, i.e., . IMU biases can be written as , where  are the gyroscope and accelerometer bias. We model them with “Brownian motion” which is integrated white noise.

Homogeneous camera calibration matrices are denoted by .  is the pose of the camera frame “” in the body frame, known from prior calibration. The “delta” pose from time  to time  is a homogeneous transformation consist by:



where we dropped the coordinate frame subscripts for readability (the notation should be unambiguous from now on).

# Chapter2 IMU 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……

# Chapter2 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……