Online Reconstruction of Indoor Scenes from RGB-D Streams

# Two-pass reconstruction scheme

### The first pass

##### Tracks camera poses at video rate and simultaneously constructs a pose graph

### The second pass

##### Takes place to handle loop closures and reconstruct the final model using globally refined camera trajectories

# Two view of point

### In computer vision, structure from motion (SFM) and **visual odometry**

### In the robotics community, closely related is what is known as (SLAM)

# Seminal works

## || Robust tracking and sparse 3D mapping in real-time||

#### 04-3DPVT- Automatic passive recovery of 3d from images and video

#### 04-CVPR- Visual odometry

#### 07-PAMI- Real-time single camera slam

#### 07-ISMAR- Parallel tracking and mapping for small ar workspaces

## || Real-time dense surface modelling||

### Outdoor scenes

###### 08-IJCV- Detailed real-time urban 3d reconstruction from video

###### 12-CVPR- Dense reconstruction on-the-fly

### Small objects and compact workspaces

###### 10-CVPR- Live dense reconstruction with a single moving camera

###### 11-ICCV- Dtam: Dense tracking and mapping in real-time

###### 13-ISMAR- Monofusion: Real-time 3d reconstruction of small scenes with a single web camera

## ||RGB-D reconstruction||

#### 11-ISMAR- Kinectfusion: Real-time dense surface mapping and tracking

#### 12-RSS- Kintinuous: Spatially extended kinectfusion

#### 12-IJRR- Rgbd mapping: Using kinect-style depth cameras for dense 3d modeling of indoor environments

#### 13-TOG- Scalable real-time volumetric surface reconstruction

#### 13-TOG- Real-time 3d reconstruction at scale using voxel hashing

#### 13-ICCV- Elastic fragments for dense scene reconstruction (QYZHOU)

#### 13-TOG- Dense scene reconstruction with points of interest

#### 14-TOG- Online structure analysis for real-time indoor scene reconstruction

#### 13-IROS- Dense visual slam for rgb-d cameras

#### 14-CVPR- Simultaneous localization and calibration: Self-calibration of consumer depth cameras

#### 15-CVPR- Robust reconstruction of indoor scenes

#### 15-IJRR- Real-time large scale dense rgbd slam with volumetric fusion

# Categorized into two groups

## || Robotics view: SLAM||

#### 11-- Real-time 3d visual slam with a hand-held rgb-d camera

#### 12-RSS- Kintinuous: Spatially extended kinectfusion

#### 12-IJRR- Rgbd mapping: Using kinect-style depth cameras for dense 3d modeling of indoor environments

#### 13-IROS- Dense visual slam for rgb-d cameras

#### 13-ICRA- Robust real-time visual odometry for dense rgb-d mapping

### Minimizing the absolute trajectory error (ATE) is the main focus

### Root mean square error (RMSE) of the ATE is used as the evaluation metric

## || Graphics and vision: reconstructed model||

#### 13-TOG- Dense scene reconstruction with points of interest

#### 13-ICCV- Elastic fragments for dense scene reconstruction (QYZHOU)

#### 14-CVPR-Simultaneous localization and calibration: Self-calibration of consumer depth cameras

#### 14-TOG- Online structure analysis for real-time indoor scene reconstruction

#### 15-CVPR- Robust reconstruction of indoor scenes

## ||Modeling indoor small objects||

#### 09-ICCV-In-hand scanning with online loop closure

#### 13-TOG- Real-time 3d reconstruction at scale using voxel hashing

## ||modeling indoor scenes||

### Real-world environments contain sophisticated structures

### Reconstructed from complex camera trajectories

### Each view covering only a small portion of the scene

# Related work

## ||High level semantic cues||

#### 12-TOG-A search-classify approach for cluttered indoor scene understanding

## ||human interaction||

#### 14-TOG-Online structure analysis for real-time indoor scene reconstruction

## ||KinectFusion||

#### 11-ISMAR- Kinectfusion: Real-time dense surface mapping and tracking

### Represents scene with a signed distance field (SDF) which is defined on a volumetric grid

### Each incoming depth image is registered to the SDF and integrated with it using a frame-to-model alignment scheme

### Real-time performance is achieved by leveraging GPGPU

## ||KinectFusion Extension||

#### 13-TOG- Scalable real-time volumetric surface reconstruction

#### 13-TOG- Real-time 3d reconstruction at scale using voxel hashing

#### 13-RSS-Realtime camera tracking and 3d reconstruction using signed distance functions

### Perform online tracking but are lack of effective mechanisms to protect against error propagation and error buildup

### In practice they cannot well handle furnished scenes that require comprehensive scanning with complex camera trajectories

# loop closure and map optimization broken up into 3 modules

## ||Frontend(pose tracker and loop detection)||

### Frame to frame

###### 12-IJRR- Rgbd mapping: Using kinect-style depth cameras for dense 3d modeling of indoor environments

###### 13-IROS- Dense visual slam for rgb-d cameras

###### 14-TOR-3-d mapping with an rgb-d camera

### Frame to model

###### 15-IJRR- Real-time large scale dense rgbd slam with volumetric fusion

### Tracks the 6DOF camera pose

## ||Backend(graph optimization and map correction)||

### Maintains a pose graph which models the geometric relations between keyframes

### Lantency is about 10 seconds for an indoor sequence that contains two looping points

### Elasticfusion recently uses incremental bundle adjustment with non-rigid deformation for real-time consistent reconstruction

##### Its point-based representation might lack general applicability compared with continuous surface representations

## ||Map generation||

#### 14-TOR-3-d mapping with an rgb-d camera

## ||Offline||

#### 13-TOG- Dense scene reconstruction with points of interest

#### 14-CVPR-Simultaneous localization and calibration: Self-calibration of consumer depth cameras

#### 15-CVPR- Robust reconstruction of indoor scenes

# Compensate for Geometric Distortion (Calibration approach)

## ||Unsurprised||

#### 13-RSS- Unsupervised intrinsic calibration of depth sensors via slam

#### 13-ICCV- Elastic fragments for dense scene reconstruction (QYZHOU)

## ||Surprised||

### LIDAR ground-truth

### RGB-D sensor test

### 3D LUT offline to compensate(80x80 image plane and discrete range

# Method

### Camera Tracking

###### 13-TOG- Real-time 3d reconstruction at scale using voxel hashing

### Pose Graph Construction

##### Detect loops with visual information and construct a pose graph incrementally

### Model Reconstruction

##### Upon termination of scanning, we perform pose graph optimization to refine the trajectory and handle loop closures

##### We make use of the volumetric fusion to fuse these depth images into a final 3D model with refined trajectory

# Preliminaries

### RGB-D video

### Camera transformation matrix

##### 

### World to camera

##### 

### Pinhole camera model with precalibrated intrinsic parameters

##### Focal lengths

##### Image centers of camera

##### Project to the image plane **under perspective projection**

3D point

##### Back projection

A pixel with depth

##### 3D world scene is represented as a TSDF (truncated signed distance function)

# Camera Tracking with Volumetric Fusion

## ||Overview||

### Adopt the framework of volumetric fusion with voxel hashing

###### 13-TOG- Real-time 3d reconstruction at scale using voxel hashing

### The scene is represented as a TSDF on a volumetric grid with voxel hashing data structure

##### Allows real-time access and update using GPGPU

##### Data can be easily streamed in and out the hash table while the camera is moving

### The volumetric fusion framework stores the TSDF valueand an associated weighting factor in a voxel whose center locates at point

##### Frame-to-model registration and model integration

# Frame-to-Model Registration

## ||Reference||

#### 13-RSS-Realtime camera tracking and 3d reconstruction using signed distance functions

#### 11-ISMAR-Kinectfusion: Real-time dense surface mapping and tracking

## ||Methods||

### Adopt a weighted TSDF tracking method

### Performs better than the original **projective ICP method**

### Registers each incoming depth image to the reconstructed TSDF

### For each pixel corresponding 3D point

### At iteration , we start to linearize around by

### Gauss-Newton

##### Linearize residuals

##### Least square

##### Denote , we get the equation

### can be computed by solve the linear equation

##### Until the change is small enough

##### Until max number of iteration(20) is reached

##### GPU parallize per-pixel computaion

13-RSS-Realtime camera tracking and 3d reconstruction using signed distance functions

# Model integration

### Apply a non-uniform weighting strategy to integrate the incoming depth image with the previously reconstructed model

##### The center of a voxel

##### Projection distance to the camera

##### The depth value on the projected pixel

##### The distance of to the underlying surface

##### Update

, get a higher weight to near ones

# Pose graph

### denote the pose graph by

### When the th frame arrives, crate a new node and add an edge connecting with predecessor

### In addition to edges that link successive frames, there are loop edges

##### To reduce data redundance and system delay, we only establish loop edges between keyframes

##### For keyframe selection, the incoming frame is identified as a new keyframe if its pose is sufficiently different from the last keyframe in the pose graph’s keyframe queue(pose change (5◦,0.02m))

##### We amend the visual place recognition technique DBoW by replacing original SURF feature with fast ORB feature

12-TOR-Bags of binary words for fast place recognition in image sequences

11-ICCV-Orb: an efficient alternative to sift or surf

##### Estimate the relative pose between two keyframes

We first extract 2D feature correspondences using a direct index strategy

Then, use RANSAC and Nister’s three-point pose algorithm [22] to calculate their relative transformation

if RANSAC inlier ratio is less than 25% or inlier number is lower than 15, we consider it as a false detection. Otherwise, we refine using all the inliers via Levenberg-Marquardt (LM) optimization

##### By now, we obtain an initial estimate of the relative pose from 2D-3D correspondences

##### ICP to refine

Instead align two 3D point clouds formed using the keyframes and their K-nearest neighbors in the RGB-D sequence

Depth maps are downsampled by half and we project 3D points to depth image to reduce the ICP search space form 3D to 2D

Since geometric-based ICP is sensitive to **planar structures**, we perform **PCA analysis** on one of the two 3D point clouds and **discard** the loop edge if points are mostly sampled from a planar object.

# Model Reconstruction

### We incrementally optimize the pose graph but not coupled with the graph construction procedure

### Control the optimization frequency to balance the delay between threads and accuracy

### Minimizing energy function

##### as a 6-vector derived from either tracking measurement or loop detection

##### as the relative transformation (also 6-vector) from pose and

##### is **Cauchy robust function**

##### is the covariance matrix of edge

### To approximate and unify the covariance from different measurements, we adopt the Monte Carlo estimation

###### 03--Multiple view geometry in computer vision

### We use the ceres-solver to solve this minimization problem

###### http: //ceres-solver.org

# Trajectory Evaluation

### RGB-D benchmark (provides synchronized ground truth camera poses by a precise motion capture system)

###### 12-IROS-A benchmark for the evaluation of rgb-d slam systems

### State-of-the-art RGB-D based SLAM systems

##### RGB-D SLAM

06-ICRA-An evaluation of the rgb-d slam system

##### MRS-MAP

13-ICCV-Largescale multi-resolution surface reconstruction from rgb-d sequences

##### DVO-SLAM

13-IROS-Dense visual slam for rgb-d cameras

##### Kintinuous SLAM

15-IJRR-Real-time large scale dense rgbd slam with volumetric fusion

### Metric: RMSE of the ATE

# Model Quality Evaluation

### Dataset

##### Sense ground truth by high precision LIDAR system (Riegl VZ 400)

##### Two indoor scenes: Reading Room and UE Lab

##### Each

A ground truth point clouds

RGB-D video stream

Corresponding calibration information

### State-of-the-art method

##### DVO-SLAM

13-IROS-Dense visual slam for rgb-d cameras

we use its optimized trajectory as inputs for volumetric fusion to obtain the reconstructed model

##### Zhou tianyi

15-CVPR- Robust reconstruction of indoor scenes

we compare with its non-rigid refinement and rigid refinement versions

##### ElasticFusion

15--Elasticfusion: Dense slam without a pose graph

We compare with the reconstructed pointbased representation

##### Error metric

Mean and median distance of the reconstructed surface to ground truth surface

14-ICRA-A benchmark for rgb-d visual odometry, 3d reconstruction and slam

For accuracy

Cumulative histogram of the distance from the ground truth surface to the reconstructed surface

For completeness

# Speed Evaluation

### On a Laptop PC with an Intel i7-4710 HQ CPU with @2.50GHz, 16GB of RAM and a nVidia Geforce GTX 980M GPU with 4GB of memory

### On the Reading Room and UE Lab datasets

### Three threads

##### Tracking and fusion threads

##### Pose graph construction

##### Optimization threads

##### Reconstruction module

### 

# 