# KinectFusion

# Framework

### Fuse all of the depth data streamed from a Kinect sensor into a single global implicit surface model of the observed scene

### Current sensor pose is simultaneously obtained by tracking the live depth frame relative to the global model using a coarse-to-fine iterative closest point (ICP) algorithm

# Two view point

## ||Computer vision: structure from motion (SfM) and multi-view stereo (MVS)||

### Camera tracking and sparse reconstructions

###### 98-ECCV-Automatic camera recovery for closed or open image sequences

### Reconstruction of dense surfaces

###### 06-CVPR-A comparison and evaluation of multi-view stereo reconstruction algorithms

## ||simultaneous localisation and mapping (SLAM)||

### MonoSLAM

###### 03-ICCV-Real-time simultaneous localization and mapping with a single camera

### Parallel Tracking and Mapping (PTAM) system

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

## ||Dense surface MVS-style||

### 

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

###### 10--Real-time dense geometry from a handheld camera

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

# Kinect Sensor

### Using an on-board ASIC a **11-bit** **640x480** depth map is generated at **30Hz**

### The depth images contain numerous ‘holes’ (No structured light depth reading was possible)

##### Certain materials or scene structures which do not reflect infra-red (IR) light

##### Very thin structures

##### Surfaces at glancing incidence angles

### When moved fast the device will also experience motion blur (like any camera) and this can also lead to missing data

# Drift-Free SLAM for AR

#### 97-IJCV-Sequential updating of projective and affine structure from motion

##### tracked camera motion incrementally, accumulating drift

#### 98-ECCV-Automatic camera recovery for closed or open image sequences

##### required off-line optimisation to close loops

#### 03-ICCV-Real-time simultaneous localisation and mapping with a single camera

##### Producing a globally consistent maps was based on probabilistic filtering of a joint state consisting of camera and scene feature position estimates

##### Targeted at small-scale workspaces compatible with some AR applications

##### In fact limited to these due the high computational cost of filtering a large state vector containing the many features that would be needed to map larger areas

##### Relatively poor due to the sparse feature maps built

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

##### Run two procedures in alternation or in parallel:

Tracking, estimating the pose of the sensor on the assumption that the current scene model is perfectly accurate

Mapping, improving and expanding the map using a form of global optimization

###### 10-ICRA-Real-time monocular SLAM: Why filter?

##### as a featurebased system, achieves excellent accuracy for camera tracking in previously unknown scenes

##### but the sparse point map it generates is still not useful for much beyond providing localisation landmarks

### live reconstruction of dense geometry

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

Using a monocular camera and dense variational optical flow matching between selected frames

able to reconstruct a patchwork of depth maps to form a dense scene model live

but relied on camera pose estimates coming from PTAM

###### 10--Real-time dense geometry from a handheld camera

Another system with many of the same ideas but further demonstrated near-real-time depth map creation

# Scan Alignment

### ICP concept

###### 92-PAMI-A method for registration of 3D shapes

poses data alignment as a nonlinear optimisation problem

Correspondences between scans are approximated using the closest pairs of points found between scans at the previous iteration

The process of obtaining the closest point correspondences is expensive

###### 92-IVC-Object modeling by registration of multiple range images

Distance metrics: Point to plane

improve convergence rates and is the preferred algorithm when surface normal measurements are available

###### 95-PAMI-Registering multiview range data to create 3D computer objects

A drastic speed up introduced by the projective data association algorithm

Available for depth data obtained in projective image form where measurements are given as a function of pixel location

###### 94-IJCV-Iterative point matching for registration of free-form curves and surfaces

Perform early iterations on a subset of possibly corresponding points

Operate within a coarse-to-fine scheme

speeding up both the data association and final pose optimization

# Dense Scene Representation

#### 87--Sensor integration for robot navigation:combining sonar and range data in a grid-based representation

##### represents space using a grid of cells

##### within each of which a probability of occupancy is accumulated via Bayesian updates every time a new range scan provides an informative observation

##### provide vital free space information together with arbitrary genus surface representation with orientation information

#### 96-SIGGRAPH-A volumetric method for building complex models from range images

##### A related non-parametric representation used in graphics is the signed distance function (SDF)

##### Fusing partial depth scans while mitigating problems related to mesh-based reconstruction algorithms

##### The SDF represents surface interfaces as zeros, free space as positive values that increase with distance from the nearest surface, and (possibly) occupied space with a similarly negative value

#### 07-CVPR-Probabilistic visibility for multi-view stereo

##### Bayesian probabilistic inference of the optimal surface reconstruction

##### Under a simple Gaussian noise model on the depth measurements with a surface visibility predicate that every surface point is visible from all sensor viewpoints

##### Results in a simple algorithm of averaging weighted signed distance function into the global frame

#### 96-SIGGRAPH-A volumetric method for building complex models from range images

##### [07] only locally true due to surface occlusions

##### truncation of the SDF as detailed in [7] is required to avoid surfaces interfering

#### 07-ICCV-A globally optimal algorithm for robust TV-L1 range image integration

##### For higher fidelity reconstructions, at the cost of extra computation, when depth maps are contaminated with heavy tailed noise or outliers

##### Uses the more robust L1 norm on the truncated SDF data term together with a total variation regularisation to obtain globally optimal surface reconstructions

### Given a SDF representation two main approaches to obtaining a view (rendering) of the surface exist

##### to extract the connected surfaces using a marching cubes type algorithm

87-SIGGRAPH-Marching cubes: A high resolution 3D surface construction algorithm

##### the surface can be directly raycast, avoiding the need to visit areas of the function that are outside the desired view frustum

98—Interactive ray tracing for isosurface rendering

attractive due to the scene complexity-independent nature of the algorithm

# Dense SLAM with Active Depth Sensing

#### 02-SIGGRAPH-Real-time 3D model acquisition

##### Combined a real-time frame-to-frame ICP implementation using the point-plane metric and projective data association

##### able to fuse depth images from the range finder for rendering purposes at rates up to 10Hz

##### lack of global pose optimisation to reduce drift prevented them from using the system for reconstructing larger scenes

##### The final models were optimized off-line using

96-SIGGRAPH-A volumetric method for building complex models from range images

##### with substantial increases in computational power, it might be possible to instead perform live volumetric SDF fusion

##### Suggest the possibility of using such an accumulated global reconstruction for resolving ambiguities that occur with frame-to-frame ICP

#### 09-3DIM-Inhand scanning with online loop closure

##### high quality scans using a fixed ToF sensor and moving object

#### 10-CVPR-3D shape scanning with a time-of-flight camera

##### Demonstrate a moving handheld ToF object scanner

#### 10-ICRA-OctoMap: A probabilistic, flexible, and compact 3D map representation for robotic systems (Octomap algorithm)

##### Relying on 3D laser range-finder data

##### Using octrees to enable scaling to large volumes

#### 10-ISER-RGB-D mapping: Using depth cameras for dense 3D modeling of indoor environments

##### estimated the live 3D motion of the sensor (1 – 2Hz update) by obtaining relativeframe alignment via ICP alignment between depth scans initialized by RGB feature matching

##### Global consistency was achieved using a pose graph optimisation by using loop closures detected using RGB feature correspondences

##### Targeted at large scale building mapping

##### The tracking front end was not designed to provide real-time performance needed for useful augmented reality

##### The dense modelling (based on a surface patch representation) provides a less refined reconstruction than can be achieved by using a full global fusion approach

# Framework

### Surface measurement

##### A pre-processing stage, where a dense vertex map and normal map pyramid are generated from the raw depth measurements obtained from the Kinect device

### Surface reconstruction update

##### Given the pose determined by tracking the depth data from a new sensor frame

##### Surface measurement is integrated into the scene model maintained with a volumetric, truncated signed distance function (TSDF) representation

### Surface prediction

##### Close the loop between mapping and localisation by tracking the live depth frame against the globally fused model

##### Performed by raycasting the signed distance function into the estimated frame to provide a dense surface prediction against which the live depth map is aligned

### Sensor pose estimation

##### Live sensor tracking is achieved using a multi-scale ICP alignment between the predicted surface and current sensor measurement

### 

# Preliminaries

### Transformation matrix

##### Euclidean group

##### Maps the camera coordinate frame at time into the global frame

##### A point

### Constant camera calibration matrix

### Projection function

### Denote homogeneous vectors

# Surface Measurement

### A raw depth map

##### at image pixel

##### is a metric point measurements in the frame

### Apply a bilateral filter to the raw depth map

###### 98-ICCV-Bilateral filtering for gray and color images

##### 

##### is a normalizing constant

### Back project the filtered depth values into vertex map

### Compute normal between neighbors

### Define a vertex validity mask

##### for a where a depth measurement transforms to a valid vertex

##### if a depth measurement is missing

### PS: The filtered version greatly increase the quality of the normal

### Multi-scale representation of vertex and normal map pyramid

##### Level

##### A depth map pyramid

is computed from by block averaging followed by sub-sampling to half the resolution

Depth values are used in the average only if they are within of the central pixel to ensure smoothing does not occur over depth boundaries

##### A normal map pyramid is computed from the corresponding depth map

##### For transform from camera to global frame

# Surface Reconstruction

## ||the volumetric truncated signed distance function (TSDF)||

#### 96-TOG-A volumetric method for building complex models from range images

### SDF

##### The value corresponds to the signed distance to the closest zero crossing (the surface interface)

##### Positive and increasing values moving from the visible surface into free space

##### Negative and decreasing values on the non-visible side

## ||TSDF||

### The global TSDF is denoted by

### Each volume consists of two components

### Rule

##### Points that are within visible space at distance greater than from the nearest surface interface are truncated to a maximum distance

##### Non-visible points farther than m from the surface are not measured

##### Otherwise the SDF represents the distance to the nearest surface point

#### 09--Remarks on the O(N) implementation of the fast marching method

## ||Projected TSDF||

### 

### Weight is portional to

### we have found that approximation within the truncation region for 100s or more fused TSDFs from multiple viewpoints (as performed here) converges towards an SDF with a pseudo-Euclidean

### Under an norm of point-wise SDF minimizing

### Incremental solution

##### Weighted running average

96-TOG-A volumetric method for building complex models from range images

### Weight truncation

# Ray casting

## ||Surface Prediction||

### Vertex map

### Normal map

## ||per pixel ray raycast||

#### 09--Remarks on the O(N) implementation of the fast marching method

### Each pixel is marched for minimum depth

##### Stop when a zero crossing from to , indicates surface

##### Stop when a zero crossing from to +, indicates back surface

##### Stop when working volume is met, indicates no measures

## ||Normal estimation||

## ||min/max block acceleration||

### Classically use to speed up marching through empty space

###### 09-- Remarks on the O(N) implementation of the fast marching method

### March along the ray in steps with size

## || Higher quality intersections||

### Can be obtained by solving a ray/trilinear cell intersection

###### 09-- Remarks on the O(N) implementation of the fast marching method

### Given a ray has been found to intersect SDF where and at points along ray and from the start point

# Sensor Pose Estimation

## ||Assumptions that allows to use all of data||

### small motion from one frame to the next

##### Allows use the fast projective data association algorithm to obtain correspondence

95-PAMI- Registering multiview range data to create 3D computer objects

02-SIGGRAPH- Real-time 3D model acquisition

##### Point-to-plane distance

92-IVC- Object modeling by registration of multiple range images

### Second, modern GPU hardware enables a fully parrallelised processing pipeline

## ||Error definition||

## ||Solution||

### Initialize

### Linearization

### Validity check and relocalisation

# Experiment

## 

### 