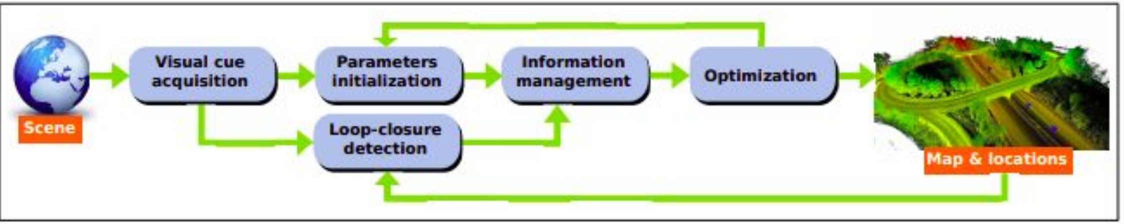
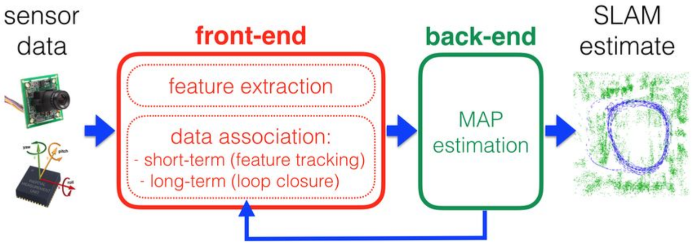
Visual SLAM

Senthil Palanisamy

High level overview of VSLAM





1. Visual cue acquisition and loop closure detection is popularly called the front end
2. Optimisation and information management is called the backend

Classification of SLAM methods based on features used

1. Feature based
2. Direct
3. Semi-direct

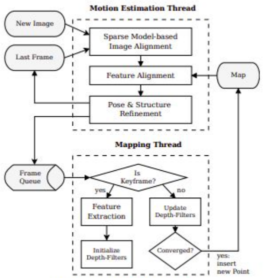


Fig. 1: Tracking and mapping pipeline

Based on density of features

1. Sparse
2. Dense

Note: A direct method is not necessarily dense. Probabilistic methods are always sparse

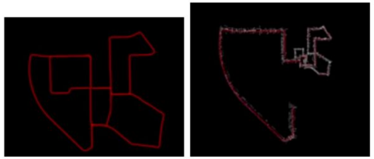
Note: Parallel thread computation is an important aspect of SLAM systems and at times, they are even the prime source of innovation in many papers [PTAM]

Classification of SLAM based core technique

1. Filter based
    a. Doesn't scale well with increase in landmarks and poses
    b. No concept of local update (The closes if FAST SLAM)
2. Optimisation based
    a. Less expensive to do local updates
    b. Global bundle adjustment is very expensive or at time impossible

Concept Glossary - Parameterisation choices

1. Camera pose:-
   a. Lie algebra and lie groups
      i. SO(3) & so(3) + 3 for x,y,z of camera pose
      ii. SE(3) & se(3)
      iii. SIM(3) & sim(3)
         1. for handling scale drift in monocular system
      iv. Quaternions
2. Inverse depth parametrization
3. What to optimise? (Variants of BA)
   a. Full BA
   b. Motion only BA
   c. Structure only BA / Pose graph optimisation
   d. Local BA
   e. Global BA
   f. 3D geometry alignment(ICP)
4. Map representation
   a. Surface (TSDF)
   b. 3D point cloud
5. Error metrics
   a. RMSE
   b. Odometry error



(a) Ground truth trajectory of a map (b) A trajectory estimated by monocular system

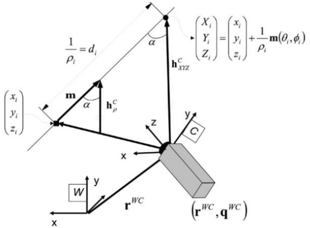


Fig. 2: Inverse depth parameterisation

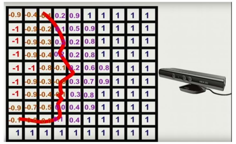


Fig. 4: A TDSF representation of the surface

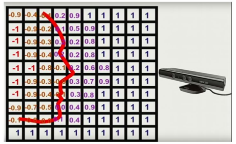
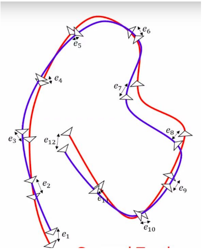
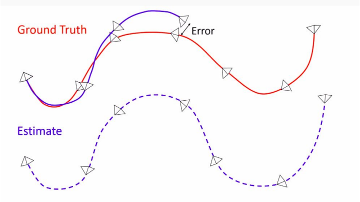


Fig. 4: A TDSF representation of the surface





Concept Glossary optimization methods and techniques

Non-linear optimisation methods

1. Gradient descent

$$x_{k+1} = x_k - \epsilon \frac{dE}{dx}(x_k)$$

2. Newton's method

$$x_{t+1} = x_t - H^{-1} g$$

3. Gauss-Newton Method

$$H_{jk} = 2 \sum_i \left(\frac{\partial r_i}{\partial x_j} \frac{\partial r_i}{\partial x_k} + r_i \frac{\partial^2 r_i}{\partial x_j \partial x_k} \right) \qquad H_{jk} \approx 2 \sum_i J_{ij} J_{ik},$$

$$x_{t+1} = x_t - (J^T J)^{-1} J^T r$$

4. Levenberg-Marquardt

$$x_{t+1} = x_t - (H + \lambda I_n)^{-1} g$$

Bundle Adjustment Parameterisation choices

1. 3D points for map

$$E(R, T, X_1, ..., X_N) = \Sigma_{j=1}^{j=N} |\tilde{x_1}^j - \pi(X_j)|^2 + |\tilde{x_2}^j - \pi(R, T, X$$

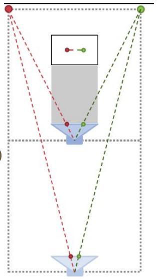
2. 2D image coordinates

$$E(x_1^j, \lambda_1^j, R, T) = \Sigma_{j=1}^{j=N} |\tilde{x_1}^j - \tilde{x}_1^j|^2 + |\tilde{x_2}^j - \pi(R \lambda_1^j x_1^j + T)|^2$$

Note: Optimisation methods 3 and 4 are commonly used in Literature

Different Sensor suites Combinations for VSLAM

1. Monocular
   a. Optimisation on temporal stereo (already discussed in class)
   b. **Specific bootstrapping:** The system should b told about the scale of the system.
      i. 5 point or 8 point algorithm
   c. Inevitable scale drift

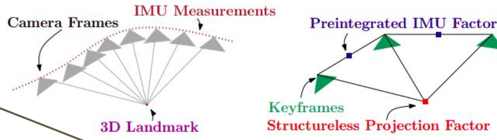


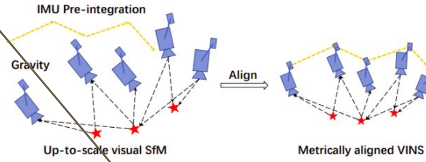
2. Stereo / Multiview
   a. Temporal + static stereo
   b. Direct point initialisation
   c. Increased cost of computation

$$E = \sum_{i \in F} \sum_{p \in P_i} \sum_{j \in obs^t(p)} E^p_{ij} + \lambda E^p_{is}$$

3. RGBD
   a. Direct depth is more accurate than inferred depth
   b. Limited range (a few meters in case of kinect xbox)
   c. Correspondence matching can be turned into a geometry matching problem (ICP in kinect fusion) : Point locations are fixed, the only learnable parameter are the camera pose that aligns the observed point cloud with the reference point cloud

4. Visual Inertial
   - Manifold Preintegration
   - Tightly coupled fusion
   - Loosely coupled fusion
   - Only 4 DoF for camera pose: 3 for location and one for orientation (No drift in roll and pitch angles.IMU gives absolute measurements)





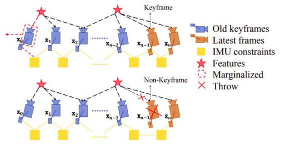
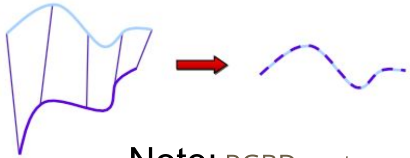


Fig. 24: An illustration of a tightly coupled VIO system



Note: RGBD systems and stereo system can be unified under a single framework.[ORB SLAM2]

$$u_R = u_L - \frac{f_x b}{d}$$

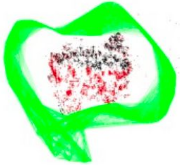
Data Management

$$\|A\theta - b\|^2 = \|Q\begin{bmatrix}R\\0\end{bmatrix}\theta - b\|^2$$

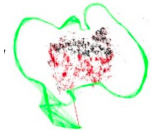
$$\begin{aligned}\|A\theta - b\|^2 &= \|Q^TQ\begin{bmatrix}R\\0\end{bmatrix}\theta - Q^Tb\|^2\\ &= \|\begin{bmatrix}R\\0\end{bmatrix} - \begin{bmatrix}d\\e\end{bmatrix}\|^2\\ &= \|R\theta - d\|^2 + \|e\|^2\end{aligned}$$

$$\begin{bmatrix}Q^T & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}A\\ w^T\end{bmatrix} = \begin{bmatrix}R\\ w^T\end{bmatrix}$$

1. Key Frames
   a. When to insert a key frame and when to delete a keyframe? Too many heuristics and mostly empirical.
   b. At its core, the real problem we are asking is “Has the scene changed enough to insert a new reference key frame” (for insertion). “Am I approaching my memory limits and which frame is the most irrelevant that can be sacrificed? (deletion)
2. Total data is represented as a graph
   a. Nodes represent the camera poses and Connection between the nodes contain the common points visible between the two views(covisibility graphs)
   b. Dense → Sparse connection: Essential graph (Essential graph is a spanning tree of the visibility graph)
   c. G20 - Open source framework
3. Windowed Optimization
   a. Local motion only BA → Local structure only BA → local BA → Global BA across key frames (Not every paper follows this but some kinds of heuristic strategy like this)



Covisibility graph

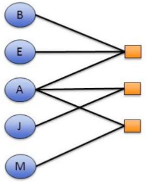


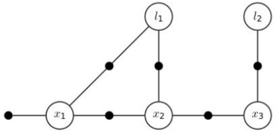
Essential graph

Factor graphs:-

Let me try a an impossible task: A 30 second intro to why factor graphs are useful

1. Bipartite graph
2. QR factorisation is fast. Once factorized, Givens rotation can be used for getting very fast results
3. The A matrix can be interpreted as a factor graph and all operations that made QR factorisation fast can be redefined on a factor graph
   a. Why? Operations are more intuitive and not abstract like abstract matrix operations
   b. Not necessarily restricted to edges - They can even contain hyperedges





$$A = \begin{array}{c} \begin{array}{ccccc} l_1 & l_2 & x_1 & x_2 & x_3 \end{array} \\ \begin{bmatrix} X & & X & & \\ X & & & X & \\ & X & & & X \\ & & X & & \\ & & X & X & \\ & & & X & X \end{bmatrix} \end{array}$$

Data association (Loop Closure)

Data association has three similar but slightly subtle problems

1. Loop closure - Have I already visited this place
2. Kidnapped robot - I am lost. Map, can tell me where I am?
3. Cooperative mapping - Has this been already mapped by another robot

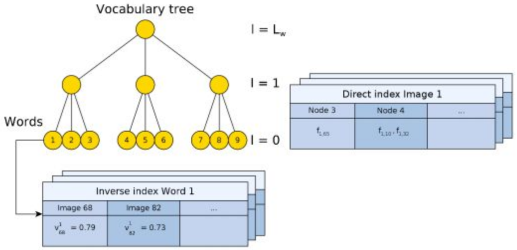
Loop closure strictly necessary properties

1. Zero False positive.
2. Non-zero true positive

Usually systems are very conservative (8-10 percent TP 0 FP).

Universal Solution to Loop closure agreed upon in the Visual SLAM community

1. Bag of words.
2. DBoW2 library



Loop closure common techniques:-

a. Image to Image (A default go to for CV scientist)
b. Map to map (A default go to AR scientist but disregarding all visual information)
c. Image to map (Shown to be promising)

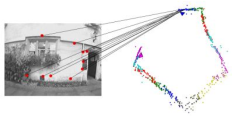


Fig. 13: image to map matching

Fig. 11: map to map matching

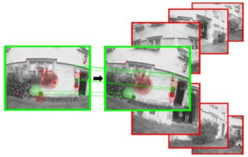


Fig. 12: image to image matching

Open question:-
Can we exploit both geometry and visual information for more efficient matching

Disconnected topics: Fast SLAM, Active SLAM, Semantic SLAM

Fast SLAM core idea:-

1. The joint distribution of camera pose and landmarks can be factored as (that why its called factorised SLAM)

$$p(s^t, \theta \,|z^t, u^t, n^t) = p(s^t|z^t, u^t, n^t)\pi_k p(\theta_k \,|s^t, z^t, u^t, n^t)$$

2. Use a particle filter for representing the conditional distribution of states $p(\theta_k \,|s^t, z^t, u^t, n^t)$
3. Use a EKF filter for representing the conditional distribution of landmarks $p(s^t|z^t, u^t, n^t)$
4. Conditional distribution is assumed between the landmarks and other previous camera poses to simplify covariance to a 2 x 2 matrix.
5. A binary search tree to speed up new estimation

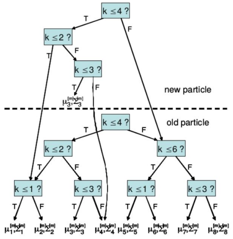


Fig. 19: A binary search tree data structure for storing the landmark estimates

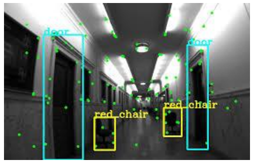
Active SLAM:-

1. This was actively followed by Davison in his series of papers under “Active Vision”
2. Invert the SLAM problem: Determine which measurements need to taken next so that the overall uncertainty in the system is minimised.
3. Purely Information theoretic view
4. Core idea:-
    a. Find the innovation covariance ellipsoid volume for each landmark.
    b. Observe the landmark with the highest uncertainty estimate.

$$V_s = \frac{4\pi}{3} n_\sigma^3 \sqrt{\lambda_1 \lambda_2 \lambda_3}$$

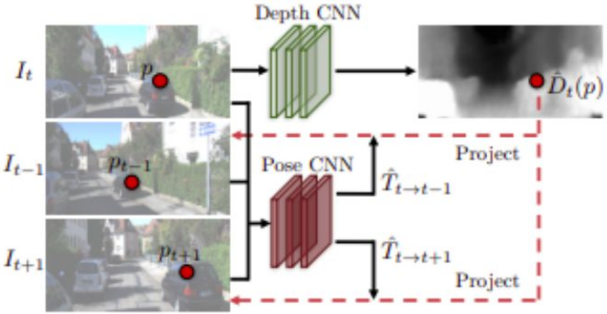
    c. The original paper idea is slightly more involved than this. This is only a distilled core idea view

Semantic SLAM:-

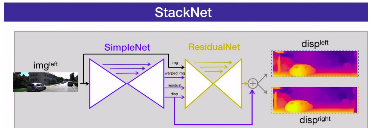


Justified use of deep learning - A personal Opinion

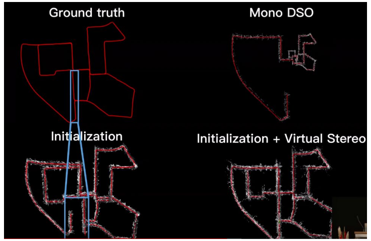
End to end deep learning

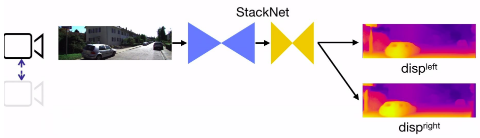


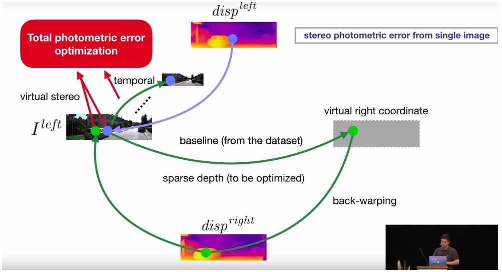
1. Learning depth for monocular images



2. Deep Virtual Stereo odometry







3. Semantic mapping

Active researchers in the area

Daniel Cremers, TUM, Germany
Optimisation, Direct methods

Andrew Davison,
Imperial College, London
Probabilistic methods, AR applications

Cyrill Stachniss, University of Bonn, Germany,
Particle Filters, Semantic SLAM, SLAM for precision robots

Sebastian Thrun,Stanford, US.
Mathematical framework of SLAM systems. (Focuses on AI, in which SLAM is one of his interests)

Michael Kaess, CMU, US.
Factor graphs, Sparisty representations,
Incremental updates

luca carlone, MIT, US
DARPA Competitions, Semantic SLAM, VI SLAM systems

John Leonard, MIT, US
SLAM for underwater robots

Frank Dellaert, Georgia tech, US,
Monto Carlo methods, Particle filters

Note: The researchers given here and the research topics given under each researcher is by no means exhaustive. These are few researchers who were source of papers during my survey and these are few areas I read a paper about them. SLAM is a pretty big area. I am sure I must have left out someone important or some topics important under some of these reseachers