Project Proposal of a Non rigid transformation based Localization and Mapping

Group 51

170387F – M. W. G. Vihan Melaka

Contribution – *Full proposal and all the references*

Supervisors

Dr Chandana Gamage
Dr. Sulochana Sooriyaarachchi

9th August 2021

2017 Intake

Department of Computer Science and Engineering
Faculty of Engineering

University of Moratuwa

Table of Content

Figures and Tables

1. Introduction

With the emergence of autonomous vehicles and improvement in the robotics field, the need of having a proper way to localize the robot of vehicles in the environment and mapping of the environment was an undeniable requirement of the field. Localization of a robot means knowing exactly where the robot is at the moment in the environment in order to achieve this robot must have some knowledge about the environment that is where the requirement of the map of the environment comes into action. Over the years scientists and researchers tried to tackle localization problem and from the results they got they got to know that localization problem is a non-converging problem if it was to solve as an independent problem. But when considering to solve localization problem together with mapping at the same time localization problem was a converging problem. This was the beginning of the Simultaneous Localization and Mapping or SLAM concept in the robotics field.

Over the years many SLAM representations emerged like EKF-SLAM which is based on extended Kalman filter and probabilistic estimation of the landmarks in the map, graph SLAM which used a graph to keep track of the environment and recently more advanced SLAM techniques which uses data from the RGB-D cameras, Lidars were came into the field and SLAM approaches based on these were collectively called as visual SLAM, SLAM methods like ORB-SLAM, RGB-D SLAM, Elastic fusion falls under this category. Although there are many solutions to this problem, every technique has its pros and cons and they are suitable for a specific scenario and conditions. Because of solutions being specific to a certain scenario there is still space for development of new SLAM techniques that will benefit an unexplored set of conditions or improvement of an existing solution to adapt more requirements is also possible.

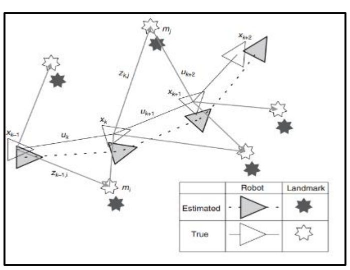

Figure 1.1 Probabilistic SLAM

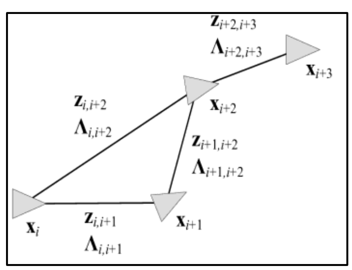

Figure 1.2 Graph SLAM

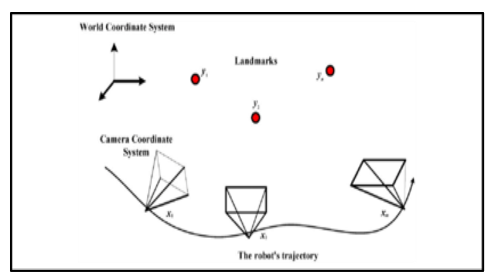

Figure 1.3 Visual SLAM

State of the art solution for visual SLAM problem consists of two parts called the front end and the backend [1]. Frontend is where all the data from the sensors is processed and localization of the robot is happening using the available data in the environment. Backend is the part where loop closures are handled. Loop closure is identifying the same location in the environment and marking that on the map as a single location. Neglecting loop closure part from the solution will simply reduce the SLAM problem to odometry. Without a loop closure map will be an infinite corridor but with the introduction of loop closure robots are able to identify loops in this corridor and they are be able to find shortcuts from one point to another without following the same path which it used to get to that point. Therefore both parts of the SLAM which is frontend and backend is equally important when creating a solution to the SLAM problem.

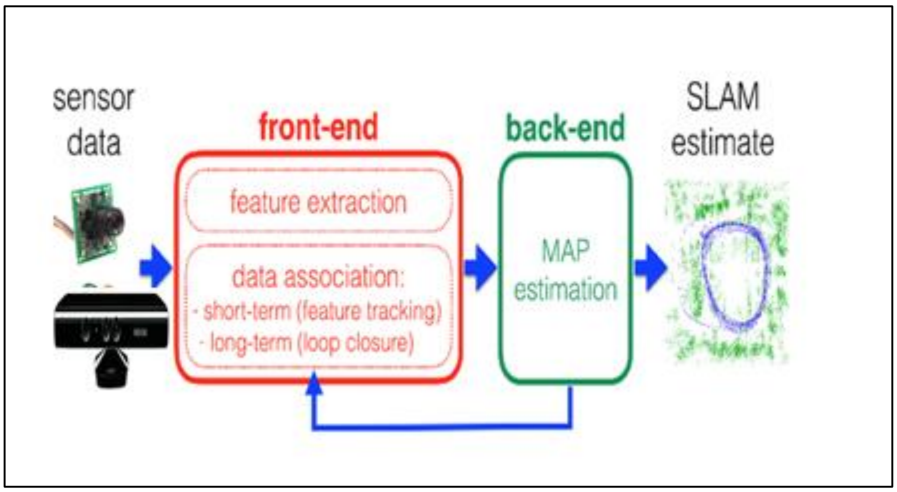

Figure 1.4 Typical model of Visual SLAM

2. Problem statement

There are two important parts in the state of the art solution for visual SLAM as frontend and the backend. For the frontend different SLAM approaches use different types of input data and different techniques of landmark extraction but in the backend there is only a limited number of techniques to use. One of the most important tasks in the backend is to reduce the error that has been accumulated in the mapping process. Most of the SLAM approaches use a technique called pose graph optimization which keeps a graph of poses of the robots as nodes and constraints like velocity, relative distance between those poses as edges of the graph representing the robots trajectory. In a loop closure situation when the backend receives two frames with identical features normally there is a small error in the pose associated with the same features, features tend to be a bit out of position because of the odometry errors. In pose graph optimization technique this error is handled by distributing that error across the pose graph as shown in the Figure 2.1. Issue with pose graph optimization technique is that because it has to maintain a pose graph, memory consumption is high, therefore it is unable to continue to map a large environment.

In a recent paper called Elastic Fusion [11] they proposed the method of using deformation theory [9] which is used in the video animation field to deform shapes, to deform the map that was build using surface features (surfels) acquired from a RGB-D camera (kinect) in a loop closure scenario as shown in the Figure2.2. But the issue with Elastic Fusion implementation was that because they were using surface features in building the map, when the number of surfels in the map increases the complexity of the system becomes too large hence mapping is limited only to map only room-sized environments. But the constructed map and the accuracy of the localization was impressive in Elastic Fusion SALM method compared to other approaches. Since Elastic Fusion SLAM approach maintain a single map and deform it memory consumption in the Elastic fusion approach is less when compared to pose graph optimization approach. In my project I am trying to improve the SLAM method used in Elastic fusion to map a larger environment by reducing the complexity involved in the computation which will result in a more reliable SLAM approach to map large environments more accurately.

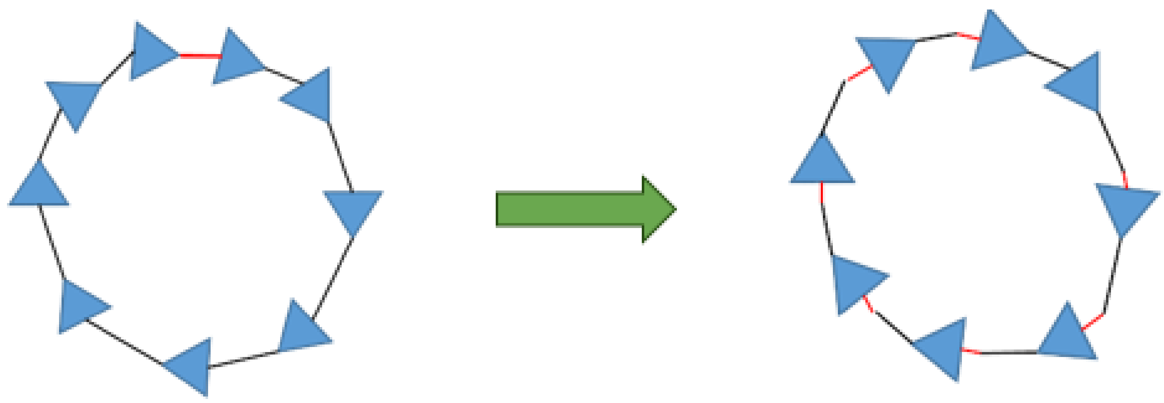

Figure 2.1 - How error at the loop closure is handled in pose graph optimization technique

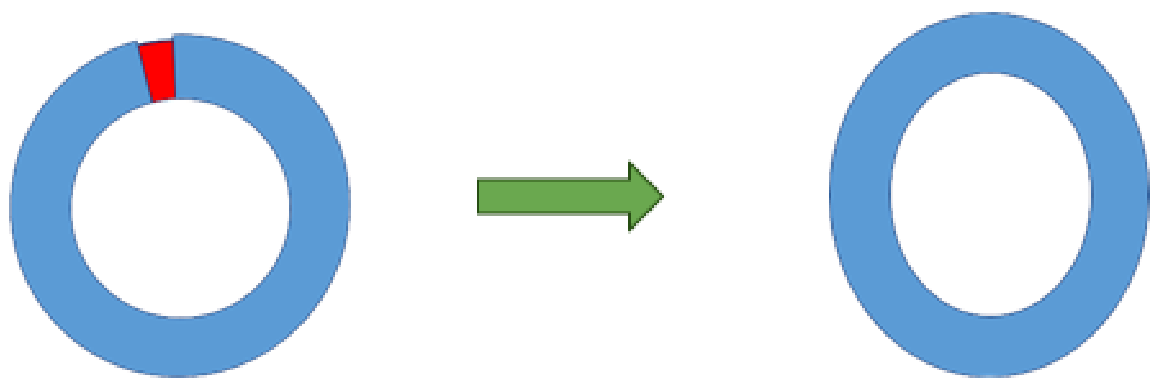

Figure 2.2 - How error at the loop closure is handled in Elastic fusion method using deformation theory

3. Research objectives

Following measures will be taken to achieve the goal of the project which is to build a SLAM system based on the work of Whelan et al [11] but with much **simpler computational complexity**.

- To reduce the complexity associated with the Elastic fusion SLAM approach [11] at the frontend a new data type will be introduced instead of surfels as the input data to the system.
- A research will be conducted on appropriate features to use in order to carry on pose estimation based on the input data type.
- Deformation theory [9] will be tested on new input data types and how to construct the deformation graph using features of new data types will also be analyzed.
- Propagating deformations from points in the deformation graph to other points that exist in the map will be carried on a GPU-based environment using CUDA.

At the end of the implementation, the proposed solution will be tested on a simulated environment built using gazebo.

4. Literature review

The most important question of, why SLAM is important in robotics field is answered in the introduction section of the work by Cadena et al on Past, Present and Future of Simultaneous Localization and Mapping. According to the work, SLAM in the robotics field is estimating robots state of a state of the robot equipped with on board sensors, and the construction of a model (the map) of the environment that the sensors are perceiving [1]. Furthermore paper shows that SLAM is not useful, for instance in an outdoor environment we can use GPS (the GPS satellites can be considered as beacons at known locations) to localize the robot and navigate in the environment. According to the work of David et al, the popularity of the SLAM problem is connected with the emergence of indoor applications of mobile robotics.

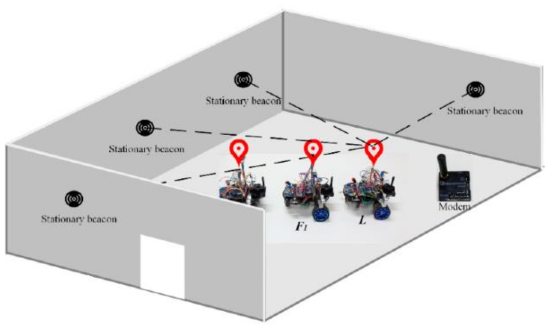

Figure 4.1 – Use of beacons for localization

Figure 4.2 – Use of GPS for localization

4.1 History of SLAM

In “The history of the SLAM problem” section of The Journal paper by H. Durrant and T. Bailey on SLAM essential algorithms [2] gives us a good overview of the initialization and evolution of the SLAM problem over the years. According to the paper at the 1986 IEEE Robotics and Automation Conference held in San Francisco first discussion about the probabilistic SLAM was discussed and as a result it was recognized that consistent probabilistic mapping was a fundamental problem in robotics with major conceptual and computational issues that need to be addressed [2]. Then paper discussed different approaches taken by researchers to solve the SLAM problem and problems faced by them, especially how the SLAM community got to know that the combined mapping and localization problem, once formulated as a single estimation problem was actually convergent [2]. From there onwards the paper gives basic ideas on probabilistic SLAM and the theory and concepts behind the probabilistic SLAM approach.

Work by Cadena et al divides the evolution of SLAM problem into two phases as classical age (1986-2004), algorithmic-analysis age (2004-2015) [1]. According to the paper classical age was the primitive stage of the SLAM problems which includes main probabilistic formulations for SLAM such as approaches based on Extended Kalman Filters, Blackwellized Particle Filters and Maximum likelihood estimation. The algorithmic-analysis age was centered on fundamental properties of SLAM, including observability, convergence and consistency. Furthermore work of Cadena et al explains what makes SLAM unique from odometry by focusing on loop closure concept and introduces modern state of the art SLAM system which consists of frontend and a backend.

At present most of the SLAM approaches fall into the category of visual SLAM. From the work of A. Huletski et al on evaluation of the modern visual SLAM methods [5] we can see that most of the modern SLAM approaches use vision frontend to capture data from the environment. The reason for the emergence of visual SLAM is because of the expensiveness of other sensors like Lidars using a monocular camera or a RGB-D camera for gathering data is much more efficient.

4.2 Related work

Project is mainly built on work of Whelan et al about the Elastic fusion SLAM approach [11]. In this paper they are integrating deformation theory which is used in the video graphics field to manipulate shapes to deform a surface in order to construct a map of the environment. This is a more map centric approach to SLAM meaning it mainly focuses on building a detailed map than having precise pose estimations. Most significant feature of this approach is how it handles loop closures using deformation theory. This implementation also uses GPU (Graphics Processing Unit) based implementation for both pose estimation and in the creation of the map. Furthermore as a continuation of this Elastic fusion method Whelan et al have also shown that this method can also be used in tasks like light source estimation as well [y].

Other than deformation theory[9] used in Elastic fusion [11] there are concepts like pose graph optimization and bundle fusion which are being used in the loop closure instances in the robotics field. Related work section of the paper by R. Mur-Artal and J. D. Tardos on ORB-SLAM2 [7] we can see that most of the RGB-D SLAM approaches use pose graph optimization in their backend. According to the paper Kintinous by Whelan et al [10], RGB-D SLAM of Endres et al. [3] and DVO-SLAM by Kerl et al [6] SLAM approaches use pose graph optimization in the backend for loop closure. But in the work of S. Rahman et al [8] on simulation of SLAM using 3D point cloud data, in the section 2.2 about Graph-SLAM they state that one of the disadvantages of SLAM that uses pose graph is that it requires high memory computations as it incorporates all the pose estimates during the calculation process which is an issue when mapping larger environment. When compared to pose graph optimization method used in Elastic fusion is more memory efficient as it only maintains a single map instead of multiple frames associated with pose.

Another method that is used in the work of R. Mur-Artal and J. D. Tardos [7] is using pose graph optimization method along with bundle fusion which is a technique used in constructing 3D environments to give the feedback of the calculated error to the pose estimation model. In the work they state that their goal is long-term and globally consistent localization instead of building the most detailed dense reconstruction which is complete opposite of the goal of work by Whelan et al. Elastic fusion by Whelan et al places much more emphasis on the accuracy of the reconstructed map over the estimated trajectory [11].

However in the related work section of the work of R. Mur-Artal and J. D. Tardos on Orb-slam2 [7], they mention about the Elastic fusion approach stating that the detailed reconstruction and localization accuracy is impressive but the problem is that the current implementation is limited to room-size maps as the complexity scales with the number of surfers in the map which is the motivation for this project. In this project a new SLAM approach which is based on work of Whelan et al on Elastic Fusion is proposed with a new data type instead of surfels which will reduce the complexity of the computation.

4.3 Deformation Theory

Deformation theory is a shape deformation technique that is mostly used in video graphics field. For the deformation of the shape this theory uses a reduce model of the shape called “Deformation graph”. Then the reduce model is deformed using affine transformations [9] [10]. Finally those transformations which were applied to the reduce model is applied to the other points in the shape in order to create the final deformed shape. Basically this process consists of four main steps.

1. Extracting key points from the initial shape to create the reduce model.
2. Building the reduce model or the deformation graph using the extracted the key points.
3. Apply affine transformations to the points in the reduce model to deform the reduce model to get the desired deformation.
4. Apply the deformations that was applied to the key points in the reduce model to the points in the shape which are not key points in order to get the final deformed shape.

This process is shown in Figure 4.3.1 using a simple example. And how this theory has used in Elastic Fusion SLAM approach is shown in the Figure 4.3.2. Top part of the image shows how deformation graph is built when an environment is mapped in a motion from left to right and then right to left and bottom part of the image shows how graph is getting built with the time.

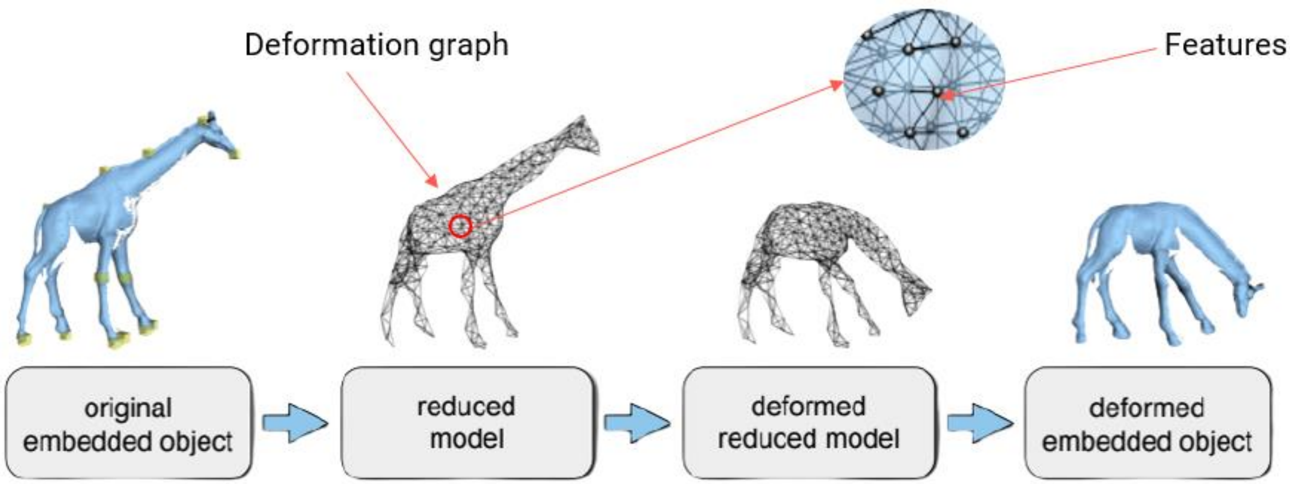

Figure 4.3.1 – Steps of Deformation theory [9]

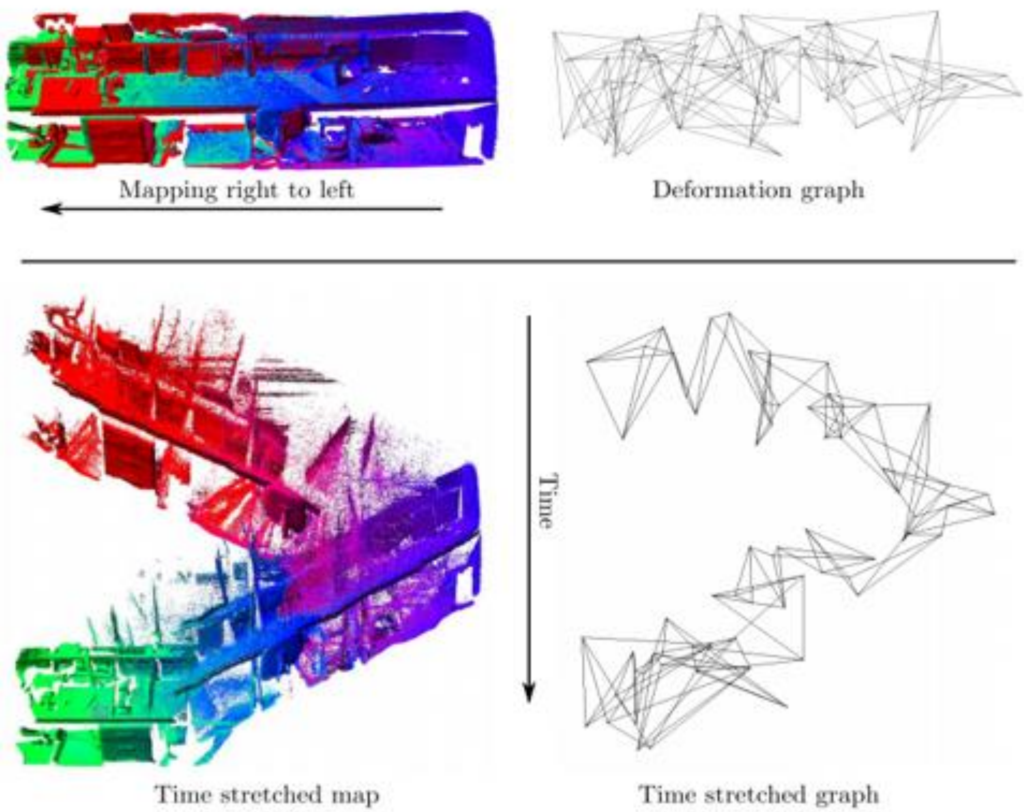

Figure 4.3.2 – Use of Deformation theory in Elastic Fusion [12]

5. Methodology

To solve the issue of high complexity of computation with scaling of Elastic Fusion by Whelan et al, in the proposed SLAM method point cloud data is proposed to be used as the input data type for the system. Same method of deformation theory [9] will be used in the backend for the loop closure. In the work of Whelan et al [11] for the pose estimation they used a GPU based approach consists of two parts called geometric pose estimation and photometric pose estimation, in the proposed approach more traditional and computationally cheap approach based on point clouds such as feature based approach where two points clouds are mapped considering their similar features and getting and idea about the motion of the robot during that period which is show in Figure 5.1 or iterative closest point where two point clouds are overlapped to get an idea about the motion of the robot during that period which is shown in Figure 5.1 is proposed to be used. Steps to follow in this type of a point cloud registration approach and some suggestions about the features that can be used is indicated in the work by Gongora Velandia [4] on Mapping of indoor environments using point cloud library.

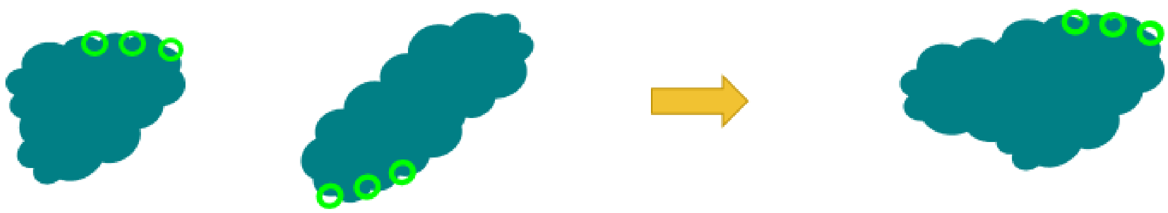

Figure 5.1 – Feature based method for localization of robot using two point clouds

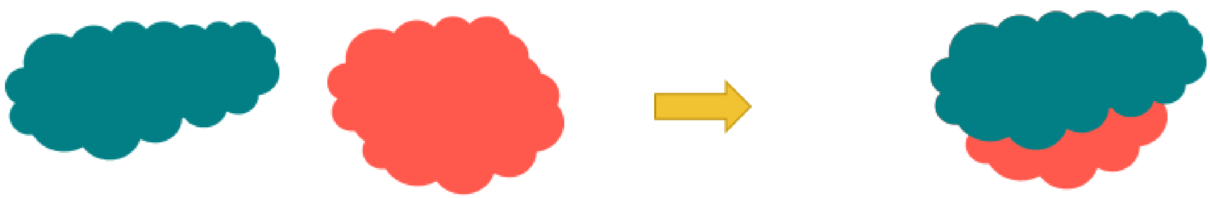

Figure 5.2 –Iterative Closest Point (ICP) method for localization of robot using two point clouds

When making the deformation graph for the mapped environment Elastic fusion [11] used most impactful surfels as the nodes of the graph, in this implementation deformation graph is proposed to be built using point cloud features. When distributing transformations to other points which are not a part of the deformation graph, Elastic Fusion uses GPU based approach and in the proposed solution the same approach will be used. Overall proposed methodology is indicated as a workflow diagram in Figure 5.3

This project is planned to carry out in a simulation environment using Gazebo and a simulated robot that runs ROS as the operating system. When working with point clouds PCL library will be used to manipulate and process data. When working with GPUs Nvidia CUDA is used as in the work of Whelan et al. [11].

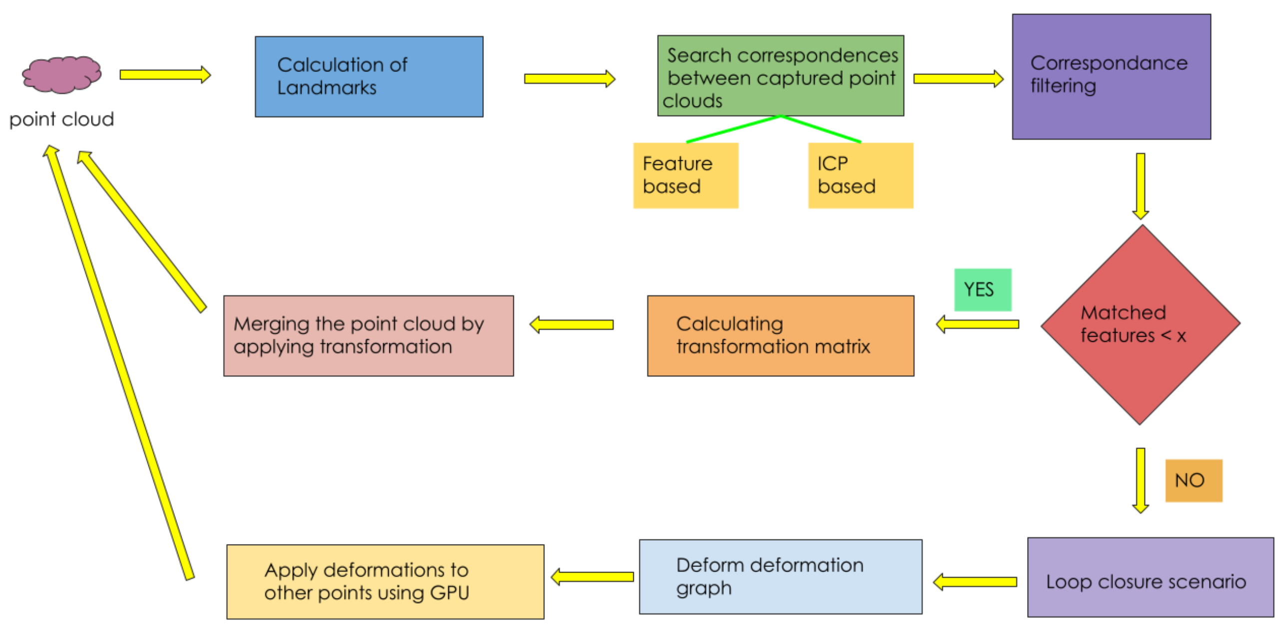

Figure 5.3 - Proposed methodology for SLAM, threshold value of "x" is yet to be calculated in an experimental manner.

6. Research Timeline

The expected project timeline for the project consists of 42 weeks of work. The Project timeline is shown in detail in Table 6.1.

Task	Week No.																																										
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	35	36	37	38	39	40	41	42
1. Literature review about SLAM																																											
2. Setting up working environment																																											
3. Studying background theories concepts and tools																																											
3.1 Studying affine transformations																																											
3.2 Studying deformation theory																																											
3.3 Studying about point clouds																																											
3.4 Getting familiar with PCL library																																											
3.5 Studying CUDA																																											
4. Implementation of proposed methodology																																											
4.1 Identifying a suitable point cloud feature																																											
4.2 Finalizing point cloud registration algorithm																																											
4.3 Trying deformation theory on point clouds																																											
4.4 Propagating deformations to other points																																											
4.5 Connecting all parts together																																											
5. Testing the implementation																																											
5.1 Building testing environment																																											
5.2 Testing the implemented solution																																											

Table 6.1 - Expected timeline for the project

7. Conclusion

In this project a new visual SLAM approach based on work by Whelan et al on Elastic fusion [11] is proposed. In the proposed solution, the current issue of complexity increasing when mapping area becomes too large with Elastic fusion approach is to be solved by using a different data type, instead of that used in the Elastic fusion method. When implementing the new solution, methods that have been used in other recent implementations of visual SLAM like RGB-D SLAM, ORB SLAM will also be taken into the concentration. Significant outlier of this approach is that, in this implementation deformation theory which is used in the video editing industry is used at the back end for loop closure which is based on the work of Whelan et al on the Elastic fusion SLAM method [11].

In this project technologies like ROS, gazebo, CUDA, PCL library will be used to achieve the expected outcomes of the project. All the development work is planned to be carried out in a simulated environment using gazebo. Testing of the new solution will also be planned to be carried out in a simulated environment. The end goal of this project is to contribute to the SLAM community with a new SLAM approach that will benefit the robotics field for its evolution towards a world-changing technology.

8. References

[1]. C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.

[2]. H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: part i," *IEEE Robotics Automation Magazine*, vol. 13, no. 2, pp. 99–110, 2006.

[3]. F. Endres, J. Hess, J. Sturm, D. Cremers, and W. Burgard, "3-d mapping with an rgb-d camera," *IEEE Transactions on Robotics*, vol. 30, pp. 177–187, 2014.

[4]. L. G´ongora Velandia, R. Hernandez Bele˜no, O. Avil´es S´anchez, and J. Rosario, "Mapping of indoor environments using point cloud library (pcl)," *International Journal of Applied Engineering Research*, vol. 11, pp. 5704–5713, 01 2016.

[5]. A. Huletski, D. Kartashov, and K. Krinkin, "Evaluation of the modern visual slam methods," in *2015 Artificial Intelligence and Natural Language and Information Extraction, Social Media and Web Search FRUCT Conference (AINL-ISMW FRUCT)*, 2015, pp. 19–25. 1.

[6]. C. Kerl, J. Sturm, and D. Cremers, "Dense visual slam for rgb-d cameras," *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2100–2106, 2013.

[7]. R. Mur-Artal and J. D. Tard´os, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[8]. S. Rahman, M. S. Abd Razak, A. Mushin, R. Hamzah, N. abu bakar, and Z. Abd Aziz, "Simulation of simultaneous localization and mapping using 3d point cloud data," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 16, p. 941, 11 2019.

[9]. S. Robert, S. Johannes, and P. Mark, "Embedded deformation for shape manipulation," *ACM Transactions on Graphics*, vol. 26, p. 07, 07 2007.

[10]. R. Sumner, J. Schmid, and M. Pauly, “Embedded deformation for shape manipulation,” in *SIGGRAPH 2007*, 2007.

[11]. T. Whelan, M. Kaess, H. Johannsson, M. Fallon, J. Leonard, and J. B. McDonald, “Real-time large-scale dense rgb-d slam with volumetric fusion*,” The International Journal of Robotics Research*, vol. 34, pp. 598 – 626, 2015.

[12]. T. Whelan, S. Leutenegger, R. F. Salas-Moreno, B. Glocker, and A. Davison, “Elasticfusion: Dense slam without a pose graph,” in *Robotics: Science and Systems*, 2015.

[13]. T. Whelan, R. F. Salas-Moreno, B. Glocker, A. Davison, and S. Leutenegger, “Elasticfusion: Real-time dense slam and light source estimation,”*The International Journal of Robotics Research*, vol. 35, pp. 1697 – 1716, 2016.