International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 5, ISSN No. 2455-2143, Pages 24-29
Published Online March – April 2016 in IJEAST (http://www.ijeast.com)

LIKELIHOOD ESTIMATION FOR SLAM USING KINECT DEVICE

Zeeshan Asim
Faculty of Industrial Management
University of Malaysia Pahang,
Lebuhraya Tun Razak 26600 Gambang
Kuantan Pahang

Dr. Shahryar Sorooshain
Faculty of Industrial Management
University of Malaysia Pahang,
Lebuhraya Tun Razak 26600
Gambang Kuantan Pahang

Ahsan Bashir
Faculty of Electronic Engineering
Sir Syed University of Engineering Technology,
Karachi, Pakistan

Abstract— **In the world of robotics, problems of visual SLAM requires an understanding of loop-closure detection and global localization, having said that in order to perform mapping and localization simultaneously we should be able to efficiently recognize an environment that has been previously visited using the current data from our RGBD camera. In this paper we present an online method to recognize and generate information regarding a previously visited place using the visual bag of words model which relies on Bayesian filtering to calculate the probability of loop closure. We would also demonstrate the robustness and effectiveness of our method by real time loop closure detection for an indoor image sequence using Microsoft Kinect camera**

Keywords—**SLAM, RGB MCL, FSLAM, EM Algorithm**

I. INTRODUCTION

In the field of robotics and specifically robotic mapping, (SLAM [1], [2]) is recognize as a computational and mathematical problem of generating and upgrading a map of a particular environment using various computational algorithms, while also should have the capacity to track the location of the agent within the map. Over the last ten years we have seen the massive development in the computer industry that aid us by providing the clients with high tech processors with better latency and instructions per cycle efficiency and it also helped in supplementing the orthodox approaches to simultaneous localization and mapping, traditional methods has their pros and cons but mostly less efficient as a consequence better algorithms has tend to be used with or replaced the bearing sensors (i.e. sonars, lasers and radio detection and ranging equipment) with single RGBD camera or camera arrays. It is important to understand that SLAM algorithms are not designed to have perfection rather compatibility with the environment they are being employed furthermore in SLAM algorithms there are issues concerning loop closure detection and global localization factors that check for the location of the agent and robot within the premises of the environment. Also in the case occlusion where abrupt Camera movements cause dissipations in path tracking the robot and keep updating the map using probabilistic approach and an online imagery database, so in this paper we have basically put forward a method to use that online imagery database and Bayesian appearance based loop closure detection to verify the 3D visual loop closure detection as the one presented in [3]. Eventually this is the problem concerning an image retrieval task to determine if the current data of image from our Kinect match the past imagery. There is a threshold value for the viewpoints of the current image and the past database, algorithm checks for the threshold value in all nearest neighbors [4] and then epipolar geometry [18] is used to detect the loop closure. In section 2, we provided a brief overview of related work on different algorithms of SLAM such as visual bag of words model and loop closure detection using Bayesian filtering, section 3 explains in few lines about the usage of visual bag of words model. In section 4 we have discussed the various filtering methods involve in the process of SLAM and the findings of our experiment are presented in section 5. The final two sections are dedicated to discussion and future enhancements

II. RELATED W0RK

The Monte Carlo Localization (MCL) method [5], recently adapted to vision ([6]) uses the map to perform global localization. The Rao-Black wellised algorithm for particle filters demonstrates loop closure probabilities to be used effectively for Fast SLAM [7], even when only bearing sensors [8] are deployed. But in all these techniques for better efficiency we need an exponential number of objects or particles are involved which is obviously intractable within the parameters of a large scale environment, which apparently leads to inefficient and abrupt sampling of data which leads to deterioration during loop closure. In this paper, we try to present a simple loop closure detection method for simultaneous localization and mapping for an online image retrieval task, following a somewhat similar approach to MCL but in a non-incremental sense. We also used the EM algorithm as presented in ([9], [10], [11]), to check for the maximum likelihood ratio for a given particle in order to perform global localization.

III. VISUAL W0RDS DICTIONARY

Bag of words model ([12], [13]) was first introduced for computer vision, it works as a sparse vector [14], which treats an image as a constituent of words, using those contents it forms an array of words and save it into the memory, when a similar image strikes the camera system matches the array of the present image with the imagery database, of course there will be a similar image in database and the system will use the image to update the existing image it helps in reducing processing time and also to perform loop closure so bag of visual words is actually a sparse vector of occurrence counts of a vocabulary of local visuals. In order to use bag of visual words model in this project, first of all an online image dictionary is built as a tree to facilitate a logarithmic time complexity in the number of words during the correlation process (description of this process is beyond the scope of this paper). In the concerned work we use SIFT key points [15] as they are famous for their robustness and reliability towards affine transformation [16], major points are marked as maxima over scale and space of differences of Gaussian convolution [17], the key points are than use to build a histogram of gradient orientations around the point which was detected earlier on the scale which is under examination.

IV. BAYESIAN LOOP-CLOSURE DETECTION

In order to understand the loop closure detection using the Bayes filtration methods it is important to first discuss briefly about Bayes probability [19] and Bayes factor ([20], [21]). In the world of probability and statistics there are basically two approaches to solve a probability problem the one is the traditional frequency approach (explanation of this method is beyond the scope of this paper) and the other is Bayesian interpretation of probability, Bayesian probability is used to solve problem where the truth and falsity of the event is uncertain. Bayesian probability is used when we have to work with the evidences in the data and foresee the result, Bayesian probability belongs to the category of evidential probabilities, and it specifies some prior probabilities in order to evaluate the hypothesis and then update the hypothesis in terms of newly calculated data or evidence.

Bayes factor can be explained as the relative evidence in the data which is also known as the likelihood ratio, Bayes factor explains the extent to which the evidence supports the concerned hypothesis, if Pr *(M)* is the probability of the given model and Pr *(D)* is the probability of the given data than mathematically the likelihood ratio can be calculated as under

$$P_r(M \mid D) = \frac{Pr(D \mid M)\,Pr(M)}{Pr(D)}$$

Where Pr (D|M) is the conditional probability for data given that probability of model is known, as a theorem Bayes factor can be defined as the posterior odds equals the prior odds times the likelihood ratio. [19] In our work we use the Bayesian probabilistic approach to ensure efficient temporal coherence (i.e. how good an image retrieves itself from the past image) and to remove any time oriented errors for detection of an image. Let S_{ti} be the arbitrary variable representing loop-closure hypotheses at time t: S_{ti} = i is the event that present image I_{time} “closes the loop” with past image I_i. This implies that the corresponding viewpoints X_{time} and X_i are close, and that I_{time} and Ii share some similarities. The event $S_{ti} = -1$ defines the event in time where no loop-closure occurred at time t. In a probabilistic Bayesian framework, the formula for the loop closure with reference to the search of the past image Ii that satisfies its index at the following equation:

$$j = \text{argmaxi}=-1\ldots t-p\ p(Sti = i|I\ time) \qquad [1]$$

Now consider if we have a series of observations O_t at a given time interval *t* we will compute the agent’s location at in a given map mt of the environment U sing the Bayes probability as follows

$$P(m_t \mid a_t, O_t) = \sum_{mt-1}^{t} P(mt \mid at, mt) \sum_{at} P(a_t \mid a_{t-1})\, P(x_t \mid m_t\ O_{1:t-1})\ /Z \qquad [2]$$

Using the above framework we can update the location posteriors of the given data sequentially such that the map of the environment and the transition function is known. Similarly we can amend the above framework to sequentially update the map of the environment as explained in the equation below:

$$P(m_t \mid a_t, O_{t-1}) \sum_{at} \sum_{mt} P(m_t \mid a_t, m_{t-1}, O_t)\, P(m_{t-1}, x_t \mid O_{1:t-1}\ m_{t-1}) \qquad [3]$$

Then by using expectation-maximization algorithm we can devise a solution for the inference problem like the one above where the solution to inferring two variables is required

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 5, ISSN No. 2455-2143, Pages 24-29
Published Online March – April 2016 in IJEAST (http://www.ijeast.com)

A. Likelihood in a Voting Scheme

Most of the bag of image models uses the extensive image to image comparison of the visual features of the data in order to update the map but on the other hand we use a technique called inverted index which is associated with the dictionary of visual words, every past image in the dictionary has been assigned a score (initially S = 0) than we use a simple voting scheme to estimate the likelihood of the image $L(S_t|Z_t)$ where zt represents the number of known words in the visual database. Such a scheme can be illustrated as follows:

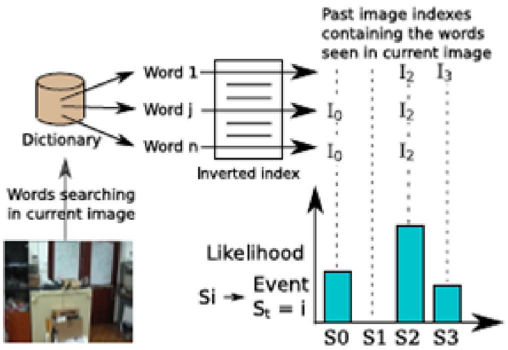

Figure 1: Voting scheme the list of the past images in which current in words have been seen is obtained from the inverted index and is used to estimate the likelihood

And finally we can use techniques of memory management [22] and inverted image modeling to localize the agent's location and to have a better recognition performance

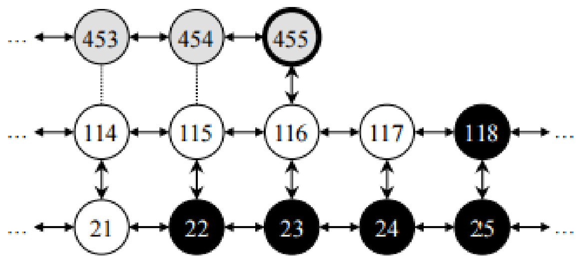

Figure 2 : Graphic representation of location vertical arrow and loop closure likes and horizontal arrows are neighbor links Dotted links shows not detected loop closure , Black location are those in LTM, white ones are in WM and Gray once are in STM Nodes 455 is the current acquired location

B. A Posteriori Hypothesis Management

At this stage we have a completely updated and stabilized posterior of the event, and after normalization of posterior we select the best candidate for loop closure by checking the probability of each posterior for the given threshold (0.85 in our case), it is obviously possible for a posterior to not have a strong peak for a unique image but rather a distribution of peaks over several neighboring images (except for I-1) due to the similarities of the neighboring visuals, so in order to obtain the right image we do not look for a single peak for a full posterior we rather look for the set of images whose sum of probabilities over the neighboring visuals exceed a set threshold keep in mind that the size of the neighboring visuals must be same. Full posterior update with the likelihood can also be understood via following diagram:

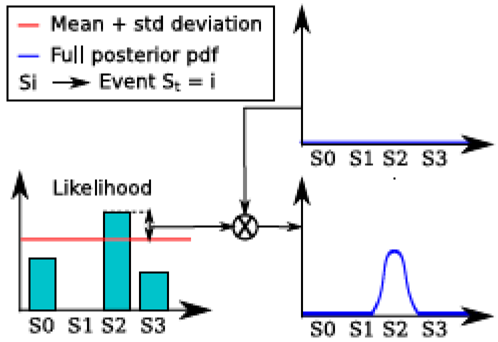

Figure 3: The full posterior pdf updated with the likelihood: when the likelihood of a hypothesis is above the mean + standard deviation threshold, the corresponding probability is updated.

V. EXPERMENTAL RESULTS

For the practical purposes of the project we used the real time appearance based mapping software as the one available at [23]. We obtained these results using a Microsoft Kinect camera interfaced with the laptop that has Kinect sdk preinstalled in it, camera process the image at 30 fps with 640 x 480 resolution. On the other hand laptop used is HP core i5 with 4 GB ram. In order to grasp the concept of this practical approach kindly refer to the example diagram below:

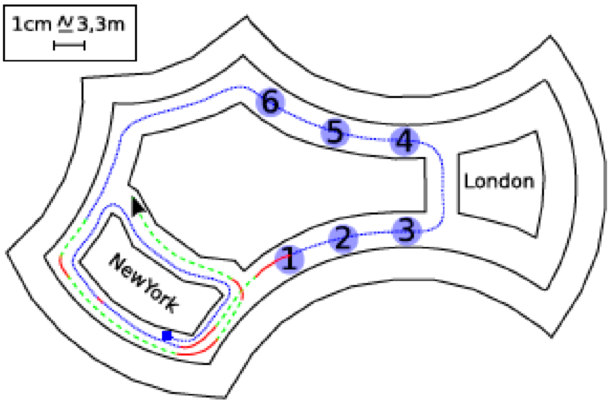

Figure4: Overall camera trajectory for the "lip6kennedy" sequence. A first loop is done around the "New York" elevators on the left before going to the "London" elevators on the right. The first loop is travelled again when the camera is back from the "London" elevators following the top-most corridor on the plan. The numbers in the blue circles indicate the positions from which the images composing the mosaic of the figure 4 were taken. See text for details about the trajectory

As we can see from the above image that the trajectory is blue every time the Kinect is rolling in undiscovered area, and as soon as it detects a loop closure it turns green and as we can see in the figure that as the camera moves in several places the green line turns red these are the places where epipolar geometry fails to satisfy the given conditions it is also known as lost odometry and it corresponds to false detection. Results from our own camera measurements are shown in the following figures

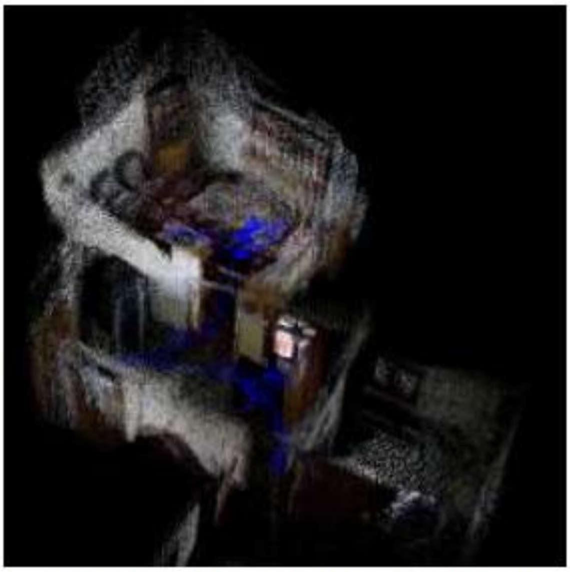

Figure 5 First Result Perceived through Camera

Blue line shows the path in which Kinect has traveled in 3-dimentional plane starting from origin

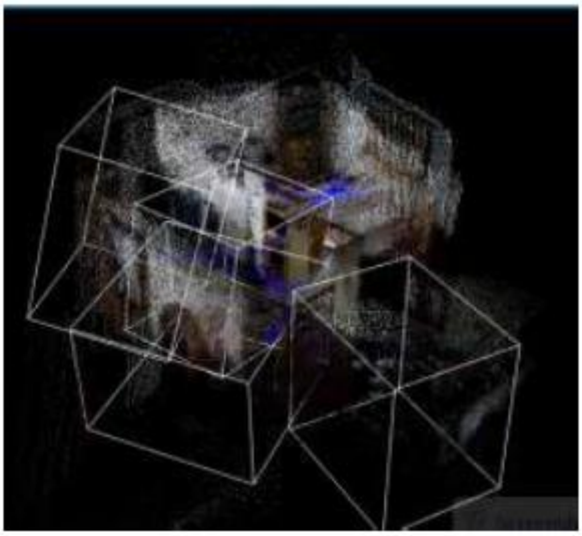

Figure 6: Add the Geometrical Dimension

Boxes show RGBD image frames used to detect odometry data and to perform optical flow field construction using Lucas-Kanade method. After completing the whole loop we then save the file with extension of .ply. This file will now be exported to MESH Lab [24] (software for 3d mapping and visualization). Important reason for extraction and export of file to Mesh Lab is that it provides a separate platform for analyzing our SLAM process and to implement the obtained map for further processes and applications. One such image from our project is shown below:

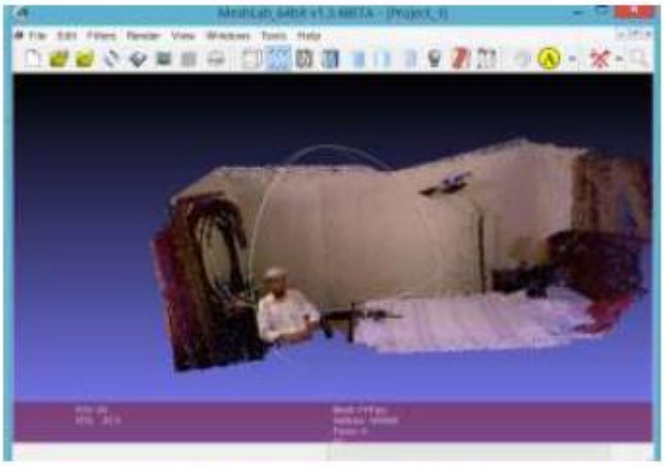

Figure 7: shows the graphical user interface of MESH Lab and the imported image from the RTAB map software

VI. FUTURE ENHANCEMENTS

As we have discussed earlier that in this paper we have tried to show a robust approach to SLAM, this is important for the practical usage of the SLAM algorithms, every projects worth is determined by its practical applications and so we will first discuss some practical applications of SLAM in general and our paper in particular various organizations such as NASA and DARPA are working on the development of autonomous robots that could be used for military and medical purposes some similar projects are discussed below

A.SELF DRIVING CARS:

Google and other automotive companies such as Volvo are working on smart cars that will have the capability to drive it autonomously, such cars use SLAM to generate the map and localize the position of the car in the area in order to keep the track of the terminal point, traffic density, and speed of the neighboring cars.

B.PLANETARY ROVERS:

Planetary rovers have the basically the same purpose as satellites sent for different planets, but unlike such artificial satellites these rovers actually have to land on the surface of the planet. In order to perform successful landing these rovers must be robust and strong, NASA and different other space research organizations make these rovers. Since these rovers need to land on the surface of the planet hence they should be equipped with proper technology to keep the track of their path and location in that unknown environment, This is where SLAM comes in very handy, using SLAM algorithms these rovers perform odometry calculations to make decisions about their particular situation.

C.UNMANNED AERIAL VEHICLES

Unmanned aerial vehicles such as drones use the SLAM methods to investigate remote areas in order to provide military with better analysis of the territory; unmanned aerial vehicles could also be used to provide medical facilities like first aid to densely populated areas and remote areas. Solution proposed in this paper is based upon visual dictionary to perform loop closure detection, which works at the detection level a better approach would be to work on the word level to perform loop closure, in future we would like to integrate our solution with the probabilistic appearance based mapping and loop closing

VII. CONCLUSION

In this paper we have put forward a robust and reliable approach to Fast SLAM in order to use visual bag of words model for perceptual aliasing of loop closure detection. We have also integrated the camera which is the Microsoft Kinect RGBD camera with the robot; this approach has helped us in understanding human machine interface and Computer vision technologies

VIII. REFERENCE

[1] H. Durant-Whyte and T. Bailey, “Simultaneous localization and mapping (slam): Part i,” IEEE Robotics and Automation Magazine, vol. 13, no. 1, pp. 99–110, 2006

[2] https://en.wikipedia.org/wiki/Simultaneous_localization_and_mapping

[3] https://spiral.imperial.ac.uk/bitstream/10044/1/1388/1/angeli_ic ra08.pdf

[4] https://en.wikipedia.org/wiki/Nearest_neighbor_search

[5] F. Dellaert, F. Dieter, W. Burgard, and S. Thrun, “Monte carlo Localization for mobile robots,” in IEEE International Conference on Robotics and Automation, May 1999

[6] J. Wolf, W. Burgard, and H. Burckhardt, “Robust vision-base Localization by combining an image retrieval system with Monte Carlo localization,” IEEE Transactions on Robotics, vol. 21, no. 2, pp. 208– 216, 2005

[7] S. Thrun, M. Montemerlo, D. Koller, B. Wegbreit, J. Nieto, and E. Nebot, “Fast SLAM: An efficient solution to the simultaneous Localization and mapping problem with unknown data association,” Journal of Machine Learning Research, vol. -, pp. –, 2004

[8] M. Pupilli and A. Calway, “Real-time visual slam with resilience to erratic motion,” in IEEE Computer Vision and Pattern Recognition, 2006

[9]https://en.wikipedia.org/wiki/Expectationmaximization_ Algorithm.

[10] J. Kosecka, F. Li, and X. Yang, "Global localization and relative positioning based on scale-invariant key points," Robotics and Autonomous Systems, vol. 52, pp. 209–228, 2005

[11] I. Ulrich and I. Nourbakhsh, "Appearance-based place recognition for topological localization," in IEEE International Conference on Robotics and Automation, 2000.

[**13]**D. Filial, "A bag of words method for interactive visual Qualitative localization and mapping," in IEEE International Conference on Robotics and Automation, 2007

[**14]**http://www.cs.umd.edu/Outreach/hsContest99/questions/node3.html

[15] D. Lowe, "Distinctive image feature from scale-invariant key point," International Journal of Computer Vision, vol. 60, no.2, pp. 91–110, 2004.

[**16**]https://en.wikipedia.org/wiki/**Affine**_transformation

[**17]**http://homepages.inf.ed.ac.uk/rbf/HIPR2/gsmooth.htm

[18]https://en.wikipedia.org/wiki/Epipolar_geometry

[**19]**https://en.wikipedia.org/wiki/Bayesian_probability

[**20]**https://en.wikipedia.org/wiki/Bayes_factor

[**21]**http://bayesfactor.blogspot.com/2014/02/the-bayesfactorpackage-this-blog-is.html

[**22]** Appearance-Based Loop Closure Detection for Online Large-Scale and Long-Term Operation Mathieu Labb´e, Student Member, IEEE, Francois Michaud, Member, IEEE

[23]https://github.com/introlab/rtabmap/wiki/Installation

[24]http://meshlab.sourceforge.net