Information Technology Journal 8 (3): 250-262, 2009
ISSN 1812-5638

Review of Feature Detection Techniques for Simultaneous Localization and Mapping and System on Chip Approach

[1]M.Y.I. Idris, [2]H. Arof, [1]E.M. Tamil, [1]N.M. Noor and [1]Z. Razak
[1]Faculty of Computer Science and Information Technology,
University of Malaya, Kuala Lumpur 50603, Malaysia
[2]Faculty of Engineering (Electrical), University of Malaya, Kuala Lumpur 50603, Malaysia

Abstract: In Vision Simultaneous Localization and Mapping (VSLAM), feature detection is used in landmark extraction and data association. It examines each pixel to find interesting part of an image that would differentiate the landmark and the less important image details. There are numerous studies in this field but they are scattered in many journals and proceedings which would require many hours just to find related material. Therefore, this research has grouped important studies done in this field to be analyzed by future researcher. Feature detection techniques such as Harris, Scale Invariant Feature Transform (SIFT), Speeded-Up Robust Features (SURF), Features from Accelerated Segment Test (FAST) and etc. is discussed in this study. A background history of each technique, their evolution and performance comparison is presented.

Key words: SIFT, SURF, FAST, SLAM, FPGA

INTRODUCTION

Localization has vast number of application such as vehicle tracking (Idna and Tamil, 2007; Idna *et al*., 2008a) and autonomous navigation. Absolute and relative measurements are two types of measurement used in establishing the location of a mobile robot. An absolute measurement device such as Global Positioning Systems (GPS) and compass depends on the infrastructure whereas relative measurement such as camera, radar and sonar are used as on-vehicle sensor. Human are more converge to the relative measurement approach in estimating their position. Based on this thought, how machine is able to use relative only measurements to calculate its own position has been a challenging subject for researchers and engineers.

Simultaneous Localization and Mapping (SLAM) is the process by which a mobile robot can build a map of an environment and at the same time use this map to compute its own location. The term SLAM was originally initiated by Hugh Durrant-Whyte and John J. Leonard at the 1995 International Symposium on Robotic Research after they and a number of researchers had been looking at applying estimation theoretic methods to mapping and localization problems (Durrant-Whyte and Bailey, 2006). One of the earliest researcher that has instigated SLAM was presented by Chatila and Laumond (1985) where, they addressed the problems which rely on defining general principles to deal with uncertainties and a methodology enabling a mobile robot to define its own reference landmarks while exploring its environment. Their approach concentrates on how to solve the problem of inaccurate sensors to model its environment and for self-location. Smith *et al.* (1988, 1990) and Smith and Cheesman (1987) discussed on the structure of the navigation area in a discrete time state space, where, a relationship between landmarks and uncertain spatial relationship is described.

AN OUTLINE OF SLAM PROCESS

SLAM process consists of series of steps that updates the mobile robot position as shown in Fig. 1. SLAM eliminates any a priori topological knowledge of the environment or artificial infrastructure to estimate the location of all landmarks (Dissanayake *et al*., 2001). The first step in eliminating this artificial infrastructure is by implementing landmark extraction and data association techniques. The extraction of different kinds of feature depends on the sensor used. Three common sensors used for feature extraction are laser, sonar and vision sensor. Most laser based range measurement employ Time of Flight (TOF) techniques and phase-shift

Corresponding Author: Mohd. Yamani Idna Idris, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur 50603, Malaysia

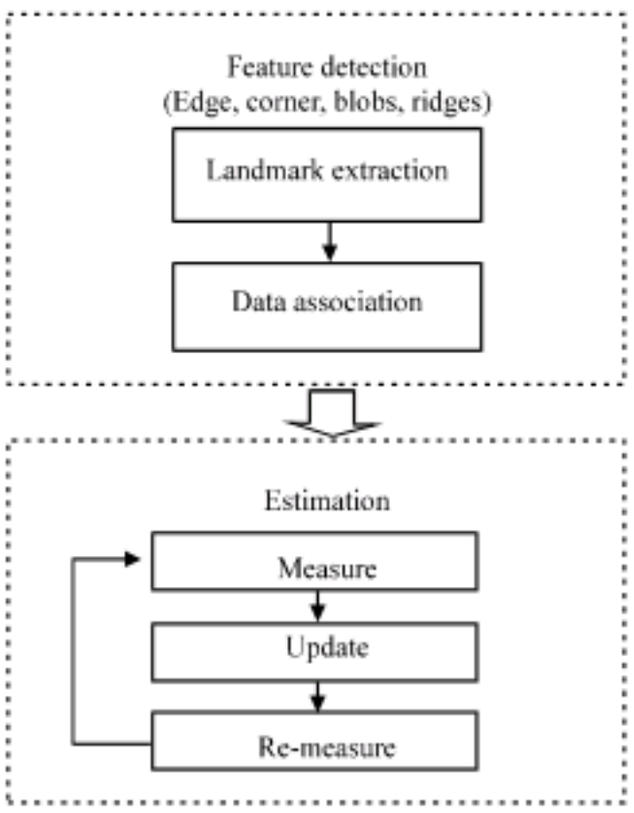

Fig. 1: SLAM outline

techniques (Zunino, 2006). Triangulation Based Fusion (TBF) (Wijk, 2001) allows the extraction of point features from the environment from sonar readings. However, among the three sensors, vision has superior capability since it is able to acquire large amount of information. With the advances in algorithms and computation power, the bottleneck data processing of a vision sensor is becoming less of a problem. Vision feature detection technique can be classified according to edge, corner, blobs and ridges detection. This feature is used to indicate the information which is specific structure in the image itself. The perceived data chosen as a landmark should be salient, easily observable and whose relative position to the robot can be estimated. The landmark must imperatively be properly associated to ensure the consistency of the robot position (Lemaire *et al*., 2007). The conventional data association metric in target tracking is the normalized innovation squared (Bar-Shalom *et al*., 2001), which is the error between predicted and actual observation normalized by error covariance.

A measurement error including slippage and time delay has introduce uncertainties in the location estimates of map landmarks. Stochastic SLAM explicitly accounts for the errors that occur in sensed measurements that show the dependency between landmark and robot pose (position and orientation) estimates. Therefore, estimation technique is used in the attempt to approximate the unknown parameters based on measured or empirical data. Most practical implementations of stochastic SLAM represent these uncertainties and correlations with a Gaussian Probability Density Function (PDF) and propagate the uncertainties using an extended Kalman filter (Nieto *et al*., 2007).

COMPUTER VISION AND SLAM

Computer vision is getting popular in SLAM landmark detection problems since vision features a high bandwidth of information if compared to laser range finder and sonar sensors. Cameras are attractive due to their lightweight and low power consumption. Issues in SLAM computer vision are much related to the problem of high detection rate and data association in recognizing previously viewed landmark and maintaining the correspondence between a measurement and a landmark (Asmar *et al*., 2006). Landmark should be easily distinguishable from their surroundings, robustly associated with the scene geometry, viewpoint invariant and seldom occluded (Davison and Murray, 2002). Problems often associated with landmark extraction are illumination or brightness variations, occlusion and perspective transformation such as scale changes and orientation translations. Others include I/O subsystem time delay to supply the processor with sufficient image data at a high speed. Many new algorithms and techniques in feature detection are proposed by researchers to overcome such problems.

Feature detection techniques: Feature detection is a low level image processing operation that represents the computation of local image features or local information content in the image. It is the starting point for landmark extraction and data association. Feature detection is closely related to interest point detection which is rich with local information content to simplify further vision system processing. The distinctive properties from the rest of the image make interest point a primary choice in vision SLAM. Feature detection can be divided into four categories namely edge, corner, blobs and ridges. Most SLAM feature detection used interest point detection technique. However, some combine more than one technique to obtain a superior result.

INTEREST POINT DETECTION

The interest points detection technique has been recognized as a popular method in visual SLAM (vSLAM) since many literatures has been published using this techniques. Point features have a clear mathematical definition with well defined position in image area. The high local information content has simplified further processing in the vision system. It is not susceptible to disturbance such as deformation (i.e., orientation or scale changes). Corner and blob detection is classified under interest point detection since it has those properties. The difference between these two techniques is only substantial when image is small. Corner looks for sharp

image features while blobs look for smooth image features. Blob detectors compliment corner detectors by detecting regions that are too smooth. Harris detection, Shi-Tomasi, SIFT, SURF and FAST are popular method categorized in interest point detection technique. Some detection though is a combination of edge and interest point detection.

Harris detection: Harris and Stephens (1988) has proposed a combined technique of corner and edge detector to cater image regions with texture and isolated feature by improving Moravac's corner detector (Frintrop *et al*., 2007). According to their research (Harris and Stephens, 1988), for an explicit tracking of image features, the image features must be discrete and not from continuum like texture or edge pixels. This is because curved lines and texture edges can be fragment differently. They used the local autocorrelation function of a signal to measure the local changes of the signal with patches shifted by a small amount in different directions. The autocorrelation matrix A, is described as:

$$A(x) = \sum_{x,y} w(x,y) \begin{bmatrix} I_x^2(x) & I_x I_y(y) \\ I_x I_y(x) & I_y^2(y) \end{bmatrix} \quad (1)$$

Where:

I_x = The respective derivative in the x-direction
I_y = The respective derivative in the y-direction
$w(x,y)$ = Weighting function (Gaussian)

$$w(x,y) = g(x,y,\sigma) = \frac{1}{2\pi\sigma} e^{\left(\frac{-x^2+y^2}{2\sigma}\right)} \quad (2)$$

The problem of Harris interest point detector is that it is not invariant to scale (Schmid *et al*., 2000). Therefore, Mikolajczyk and Schmid (2001, 2002, 2004, 2005) combines the Harris detector with the Laplacian based scale selection (i.e., Harris-Laplace) to cater the not invariant problem. The Harris-Laplace detector is then extended to deal with significant affine transformations or called Harris-affine (Mikolajczyk and Schmid, 2002). An affine transformation carries finite points to finite points and parallel lines to parallel lines but the distance between points and angle between lines can be altered. Their algorithm simultaneously adapts location, scale and shape of a point neighborhood to obtain affine invariant points (Fig. 2).

Harris corner detection has been implemented in SLAM research by many researchers. Hygounenc *et al.* (2004) in the autonomous blimp project of LAAS-CNRS used Harris detector to match visual landmark. Harris is chosen since it has good stability properties and its

Fig. 2: Mikolajczyk and Schmid (2002) correctly matched images with scale changes of 1.8 and viewpoint changes of 30°

repeatability is high enough to allow robust matches. The high repeatability shows the reliability to find the same interests point under different views. In (Royer *et al*., 2007) corner response *R* is computed as in Harris study and local maxima of *R* are potential interest points. Zero Normalized Cross Correlation over a 11x11 pixel window is used to select and discard potential matches. Frintrop *et al*. (2007) used Harris-Laplace corner detection along with visual object detection with a computational attention system (VOCUS) (Frintrop, 2006) to find region of interest (ROI). Their purpose of combining the two methods is to make the matching of region for loop closing more stable. This is feasible since the attentional ROI focus the processing on salient image regions which are thereby well re-detectable and Harris-Laplace corners provide well-localized points which enable precise depth estimation when performing structure from motion. Lemaire *et al*. (2007) rely on Harris interest point to ensure robust and reliable segment matches where segment matches are established according to a hypotheses generation and confirmation paradigm. The process has result in the widely differing observation from the rest of the data even for large viewpoint changes. This is very useful for loop closing problem where the robot able to determine whether it has reached an area that has been visited before (Fig. 3).

Shi and Tomasi detection: Tomasi and Kanade (1991), shows how to observe the quality of image features during tracking by using a measure of feature dissimilarity that quantifies the change of appearance of a feature between the first and the current frame. They also provide experimental proof that pure translation is an insufficient model for image motion when measuring dissimilarity but affine image changes that is linear warping and translation are adequate. They propose Newton Raphson minimization procedure to resolve affine changes. Region with rich texture might have problem of depth discontinuity, illumination, occlusion and drift away from

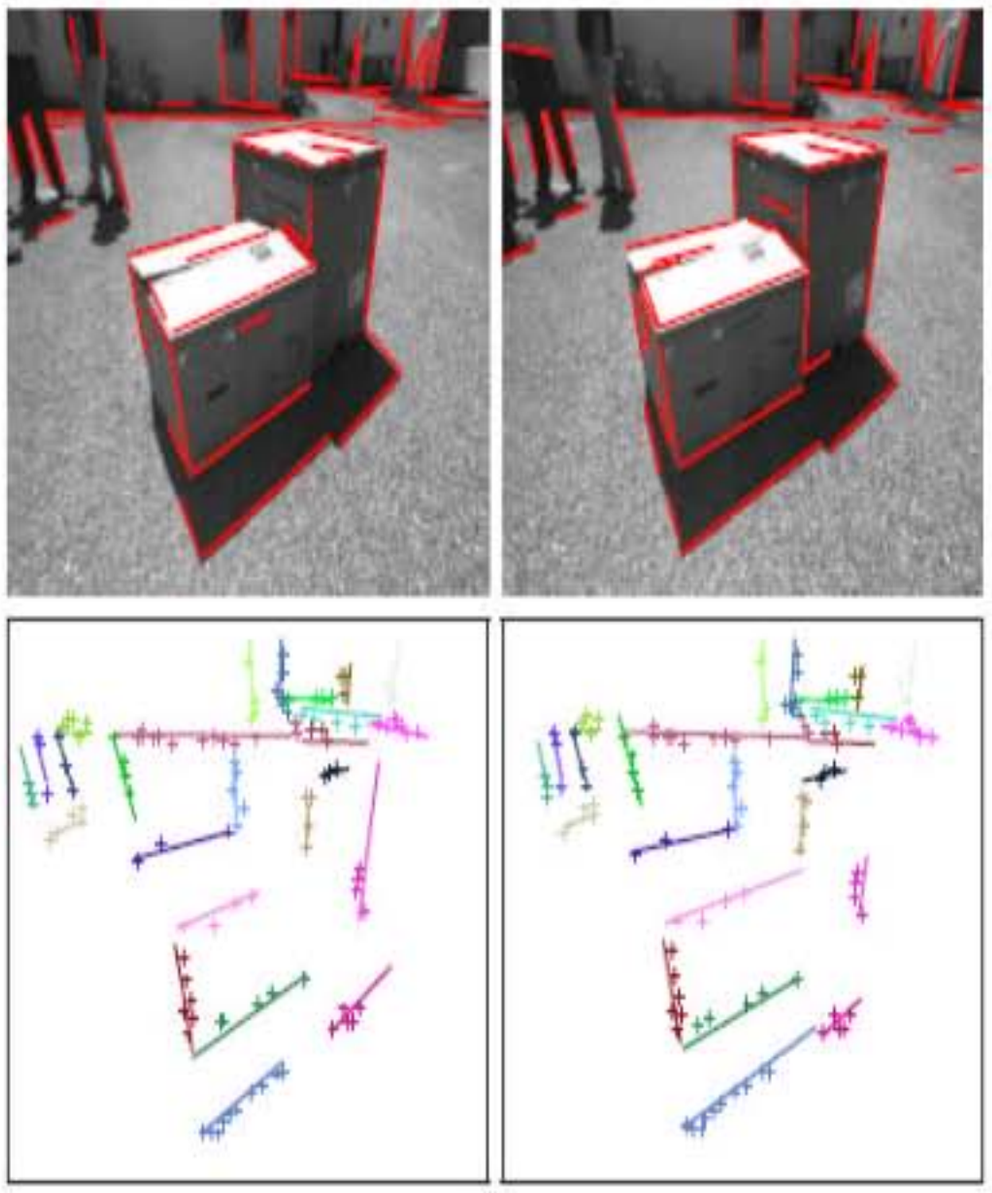

Fig. 3: Lemaire *et al.* (2007) segment matches with corresponding interest point. Matching segments are drawn with the same color

their original target. Therefore, features with good texture properties are chosen to ensure the tracker work best.

Davison and Murray (2002) and Davison (2003) used Shi-Tomasi techniques to detect interest region and match subsequent frames using normalized sum-of-squared difference correlation. Their study is extended by Clemente *et al.* (2007) which combine Shi-Tomasi with application of a Gaussian weighted window to the Hessian matrix to get more salient and better tractable features. This is to make the response of the detector isotropic and results in patches better centered around the corner or salient point.

Williams *et al.* (2007) also extend Davison (2003) studied by running Shi-Tomasi across the entire image and then performed exhaustive correlation between each of the corner points and the image patches for the whole map. A number of potential matches to map features are found in the image. However, there are features that were not detected. Williams *et al.* (2007) implemented the three-point-pose algorithm (Haralick *et al.*, 1994) to determine the camera pose indicated by each set of matches. Figure 4 results obtained by Williams *et al.* (2007).

Besides normal robot navigation problem, Shi-Tomasi is also used in Minimal Invasive Surgery (MIS) and wearable computing. The challenge in MIS is the problem

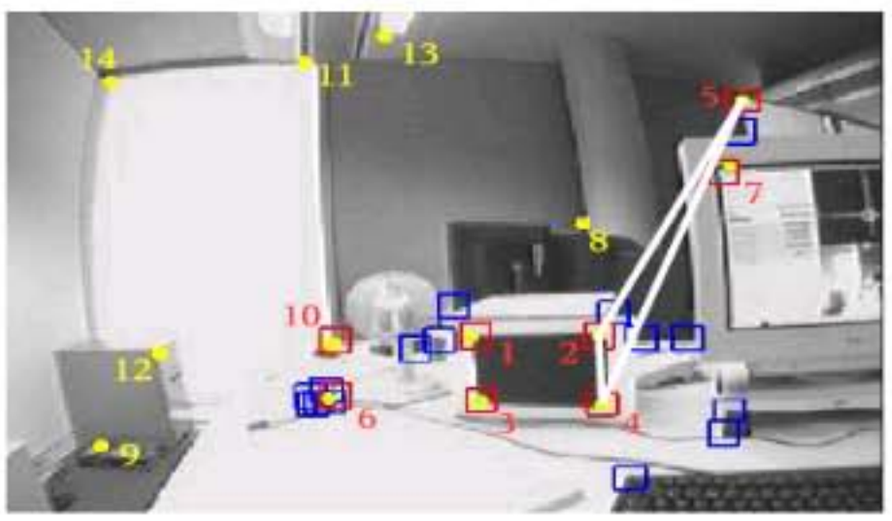

Fig. 4: Results of Williams *et al.* (2007) three point pose algorithm that finds up to four valid poses for the camera given the position in 3D

of the restricted vision can make navigation and localization within the human body. Mountney *et al.* (2006) presents a robust technique for building a repeatable long term 3D map of the scene at the same time as recovering the camera movement based on Simultaneous Localization and Mapping (SLAM) as shown in Fig. 5. Castle *et al.* (2007) detected potential insertion into 3D map with Shi-Tomasi saliency operator for their simultaneous recognition, localization and mapping for hand-held and wearable cameras.

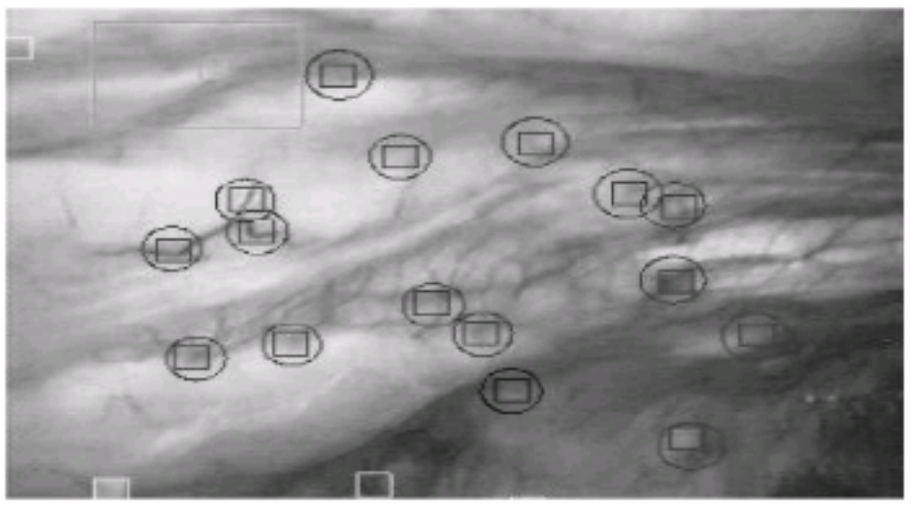

Fig. 5: A MIS scene by Mountney *et al*. (2006) where, the pixel box represent feature patches detected using the Shi-Tomasi operator

Scale-Invariant Feature Transform (SIFT): The Scale-Invariant Feature Transform or SIFT is introduced by David (1999) to improve on earlier approaches by transforms an image into a large collection of local feature vectors that being largely invariant to changes in scale, illumination and local affine distortions. A 2D Gaussian function is computed by convolving input image with two passes of the 1D Gaussian function in the horizontal and vertical directions:

$$G(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-x^2/2\sigma^2} \quad (3)$$

Lowe exploited locations that are maxima or minima of a Difference-of-Gaussian (DoG) function applied in scale space to generate local feature vector that represent an image as a one parameter. The key locations are computed by building an image pyramid with re-sampling between each level. DoG image is given by:

$$\begin{aligned} D(x,y,\sigma) &= (G(x,y,k\sigma) - G(x,y,\sigma))*I(x,y) \\ &= L(x,y,k\sigma) - L(x,y,\sigma) \end{aligned} \quad (4)$$

Where:
$L(x,y,\sigma)$ = Scale space of an image
$G(x,y,\sigma)$ = Gaussian blur
$I(x,y)$ = Original image
k = Multiplicative factor

The major stages of computation used to generate the set of image features are (David, 2004):

- Scale-space extrema detection
- Keypoint localization
- Orientation assignment
- Keypoint descriptor

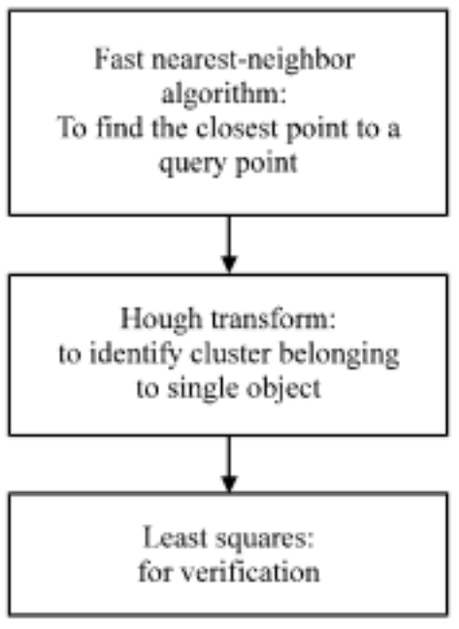

Fig. 6: SIFT object recognition process

Table 1: Averaged error between the robot actual position and the position estimated (Park *et al*., 2007)

Method	X (mm)	Z (mm)	θ (degree)
ODO-LOC	277.17	880.19	19.47
ODO-SIFT-KAL	107.37	558.70	10.34
HARI-LOC	174.30	758.83	8.85
HARI-SIFT-KAL	127.01	218.51	8.87

David (2004) modified his approach to just locate each keypoint at the location and scale of the candidate keypoint (David, 1999) to the calculation of the interpolated location of the maximum using quadratic Taylor expansion of DoG scale space function and Best-Bin-First (BBF) algorithm (Beis and Lowe, 1997). This improves the matching and stability of the selected image by approximating the closet neighbor with high probability. The matching process is done according to the sequence in Fig. 6.

Research to improve on the SIFT algorithm for SLAM problem has been done by several researchers. Park *et al*. (2007) proposed autonomous semantic-map building using their so called HARI-SIFT-KAL approach which combines Harris, SIFT and kalman filter to solve localization and kidnapped robot problem. The kidnapped robot problem is a problem refers to the case where robot is lifted and manually repositioned in a different location in the environment and has to relocate itself based on new sensor evidence. Table 1 shows their average error comparison result where map which is built by relative localization by odometer (ODO-LOC), by extended Kalman filter using odometer and 3D SIFT features (ODO-SIFT-KAL), by relative localization by optical flows of 3D Harris corners (HARI-LOC).

Schleicher *et al*. (2007a, b) put forward an idea of SIFT finger print that divide the global map into local sub-maps to achieve large closing loops in robot path running in real time. A fingerprint characterize the visual appearance of an image from some SIFT visual landmarks and the relation among them. The comparison between

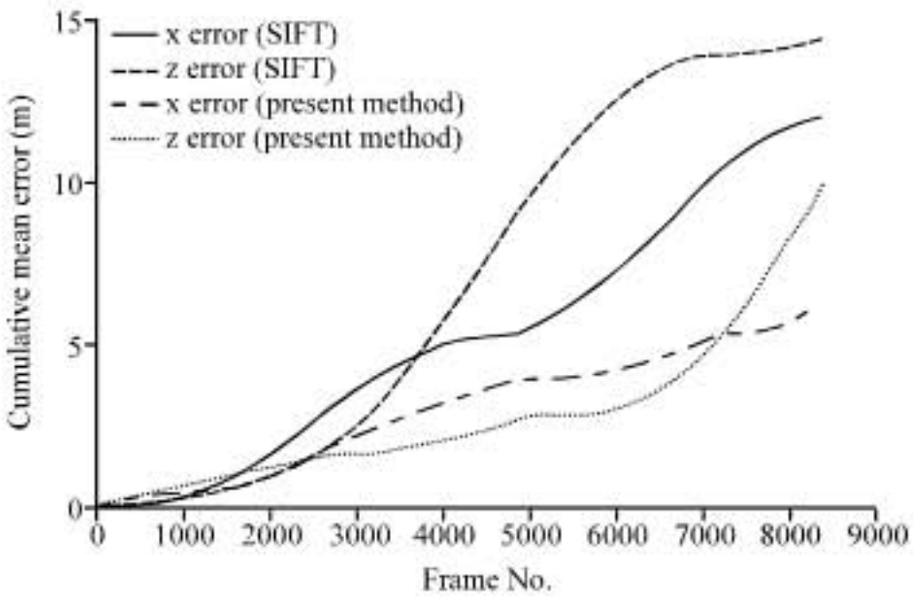

Fig. 7: Cumulative mean error for X and Z axis respect to the frame number n using SIFT implementation and Schleicher *et al.* (2007a, b) approach

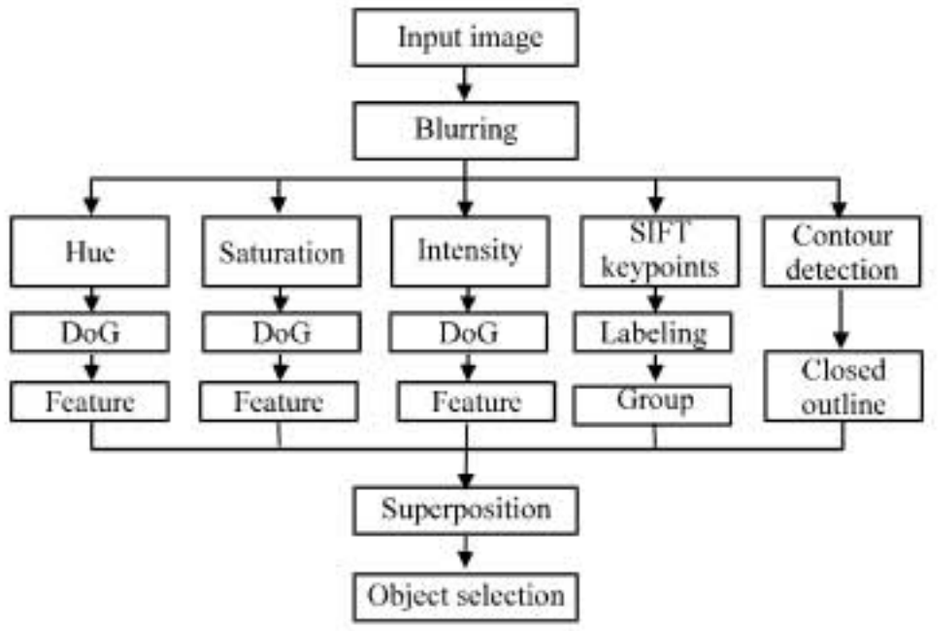

Fig. 8: Overall structure of Yong-Ju Lee and Jae-Bok (2007) proposed scheme

earlier and current fingerprint is used to detect pre-visited zones. Figure 7 shows lower cumulative mean error is obtained using SIFT fingerprints method.

SIFT keypoints single-handedly suffer from inconsistency result and would not be a reliable approach in indoor environments. For that reason, an approach is proposed by Yong-Ju and Jae-Bok (2007). The proposed image analysis consists of extracting various kinds of features and summing them. Without any prior object information, a suitable object is extracted. The detected objects are decoupled from the source image and registered in the database and used as landmarks to aid in estimating the robot pose. The overall structure of Yong-Ju Lee and Jae-Bok (2007) approach is shown in Fig. 8.

Speeded up Robust Features (SURF): Bay *et al.* (2008) presented an approach called SURF in their research. Their approach is based on Hessian matrix (Fast Hessian) and sums of 2D Haar wavelet response. They rely on the determinant of the Hessian for selecting both location and scale. A box filter is used to approximate second order Gaussian derivative. Bay also did comparison of the average recognition with selected interest point approach (Mikolajczyk's GLOH (2005); David (2004) SIFT; Ke and Sukthankar's (2004) PCA SIFT in Table 2. U-SURF is not invariant to image rotation, SURF-36 is a short descriptor with 3*x*3 sub-regions and SURF-128 is the extended descriptor for 4x4 sub-regions.

Strasdat *et al.* (2007) chose rotational dependent version of SURF in their visual bearing-only SLAM since the roll angle of the camera is fixed when it is attached to a wheeled robot. In the study Vision based SLAM for Robot Navigation with Single Camera, Shen and Liu (2007) select SURF feature points to match between image pairs.

Fig. 9: Example of SURF matching results between 2 consecutive images by Shen and Liu (2007)

Table 2: Performance evaluation comparison by Bay *et al.* (2008)

Approach	Recognition rate (%)	Time (m sec^{-1})
SURF-128	85.7	391
U-SURF	83.8	255
SURF	82.6	354
GLOH	78.3	N/A
SIFT	78.1	1036
PCA-SIFT	72.3	N/A

An example of their result is shown in Fig. 9. SURF is considered as robust to image scale and illumination changes.

Mirisola *et al.* (2007) proposed a method to recover a trajectory and 3D mapping from monocular aerial images by exploiting the inertial orientation measurements. Pixel correspondences over pairs of gray level images are established from SURF algorithm to compensate lens distortion. The problem of learning 3D maps of the environment using low cost setup (standard web cams and low cost Inertial Measurement Unit (IMU) is addressed by Steder *et al.* (2007). They combine SURF features with PROSAC-based technique (Chum and Matas, 2005) to identify the correct correspondences between images. SURF was chosen since it is comparably faster than SIFT. Furthermore, a SURF feature is rotation and scale invariant and is described by a descriptor vector and the position, orientation and scale in the image.

Table 3: Advantage and disadvantage of point features and line features

Advantages	Disadvantages
Point features	**Point features**
• Easy to localize	• Appearance of a feature point can change substantially over several frames
• Easy to find correspondences between frames	• Features which are invariant to scaled Euclidean or affine transformations can be expensive to compute
• Robust to large, unpredictable inter-frame motions	• Invariance reduce the ability to discriminate
• Larger numbers of feature points can be used to assist pose estimation while retaining real-time performance by using surface model	• Errors in the position of the surface create 3D point position errors which lead to an amplification of pose drift
• Surface model coupled with keyframes is able to provide a global database of the 3D position and appearance of feature points	• Populating a large immersive 3D environment with a sufficient number of key frames would be both difficult and very time consuming
• Good discrimination	• View dependant
Line features	**Line features**
• Stable under a very wide range of lighting conditions and aspect changes	• Difficult to find correspondences
• Robust to viewing changes	• Difficult to track edges robustly
	• More fragile

Features From Accelerated Segment Test (FAST): A feature from Accelerated Segment Test (FAST) was introduced by Rosten and Drummond (2005, 2006). They combine edge and point based tracking systems to address the problem of real-time 3D model-based tracking systems. The edge and point based complements each other and would create a more robust system. The advantage and disadvantage of both point features and line features are shown in Table 3.

To build a corner detector, a circle of sixteen pixels around the corner candidate is considered. The test is performed on a Bresenham circle. These pixels are then classified as below:

$$Sp \to x = \begin{cases} d, Ip \to x \le Ip - t & (darker) \\ s, Ip - t < Ip \to x < Ip + t & (similar) \\ b, Ip + t \le Ip \to x & (brighter) \end{cases} \quad (5)$$

Where:

$p \to x$ = Position relative to x
I_p = Candidate pixel
t = Threshold

Entropy of the positive and negative corner classification based on the pixel which yields the most information is measured using ID3 algorithm (Quinlan, 1986) to decide whether it is a corner. Non-maxima suppression is then applied on the sum of the absolute difference between the pixels in the circle and the center pixel (Tuytelaars and Mikolajczyk, 2008). The comparison between FAST approaches with other feature detection techniques is shown in Fig. 10.

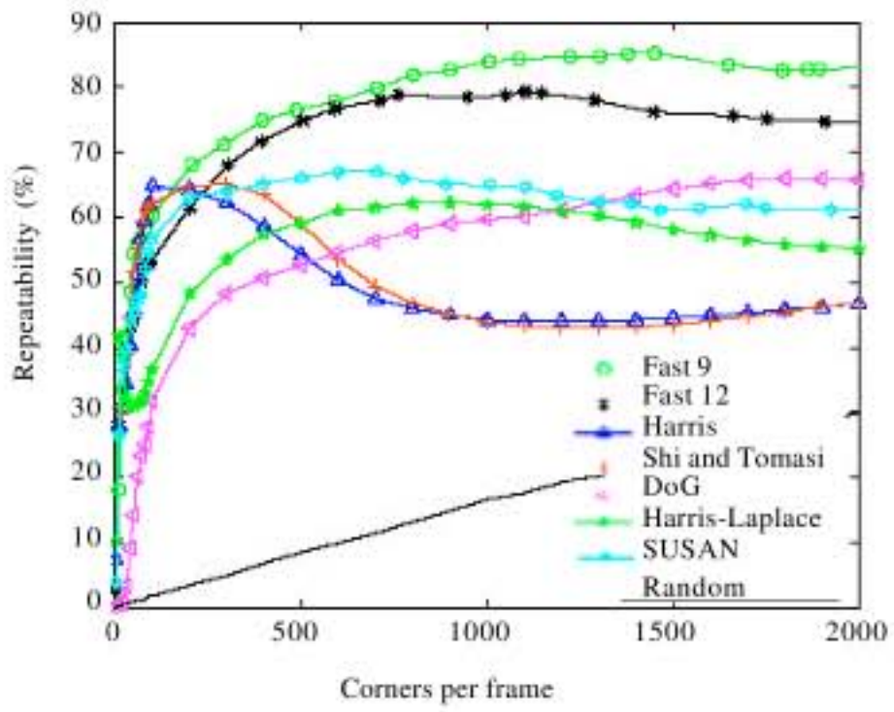

Fig. 10: Repeatability result with box datasets (Rosten and Drummond, 2006)

EDGE DETECTION

Edge detection technique mark the points in a digital image at which the image brightness changes sharply. The property changes are corresponds to several factors such as discontinuities in depth and surface or could be initiated by illumination variations and changes in material properties. Even though interest point techniques have been chosen widely by SLAM researchers, edge detection methods are chosen to compliment the drawbacks. The well known approaches in edge detection are Canny (1986) and Sobel edge detector. Most edge detection methods can be classified into two categories: search-based and zero- crossing based. Canny (1986) was categorized as zero-crossing based since it detects the zero-crossings of the second directional derivative of the smoothed image in the direction of the gradient where the gradient magnitude of the smoothed image being greater than some threshold depending on image statistics. Canny's approach is then improved by Deriche (1987) who introduced a Canny-Diriche method which use Infinite Impulse Response (IIR). IIR is recursive and can be computed in a short, fixed amount of time for any

desired amount of smoothing. Sobel operator on the other hand is used to find the approximate absolute gradient magnitude at each point in an input grayscale by using simple convolution kernel:

$$N(x,y) = \sum_{k=-l}^{l} \sum_{j=-l}^{l} K(j,k)p(x-j,y-k) \quad (6)$$

Canny edge detector: Eade and Drammond (2006), a well localized edge landmark and efficient algorithm for selecting the landmark are presented. Edge features are chosen because it contains higher-order geometric information that useful both during and after SLAM. Canny's edge detector is used as the starting point of their edge feature selection algorithm. Their study is concerned with estimating edges of arbitrary location and orientation. Asmar *et al*. (2006) also used Canny's algorithm in their SLAM solution to find dominant edges in the vertical direction of a tree trunk. The tree trunk is used as a landmark for outdoor cluttered environment. They modified Canny's algorithm to increase the traceability of the tree profile by giving more weight to the vertical (Fig. 11). Other researchers who implement Canny edge detector in SLAM are Fu *et al*. (2007), Escolano *et al*. (2007) and Taylor *et al*. (2007).

Sobel edge detector: Sobel edge detector has been used by Shaw and Barnes (2006). They proposed a detector that finds perspective rectangle structural features that would recover the edge point that are aligned along vanishing lines. Sobel edge detector blurs edge features which then provide a candidate line to be more readily detected. In Harati *et al*. (2007) suggest planar surface are more suitable than point clouds raw data since points are too redundant to be directly used in mapping for indoor environment. They use 1D orientation measure called Bearing Angle (BA). They claimed that it is more efficient to use edge based approaches and BA-based segmentation in navigation applications. A closing with a plus shape structuring element (3x3 pixels) with Sobel edge detector is used in their experiments. Smith *et al*. (2006) claim that Canny and Hough transform are too slow for real time operation. In monocular SLAM, a line detector is simply to find enough good line in a frame that sufficient can be initialized for tracking and not need to be complete or repeatable. In the preprocessing stage, Smith *et al*. (2006) make use of FAST and Sobel algorithm. FAST is used to identify corner feature while Sobel is used to test possible lines. Then the test step in Fig. 12 is taken (for 320x240 images).

To ensure the map remain sparse, the line near to existing line is ignored using the post processing stage by checking the overlapping lines. This renders each line into a Boolean image, starting with the existing lines from the SLAM system and then rendering the new lines in order of length, starting with the longest first since longer lines is more preferred. Any line which overlaps an already-rendered line is rejected. Figure 13 shows the Smith result.

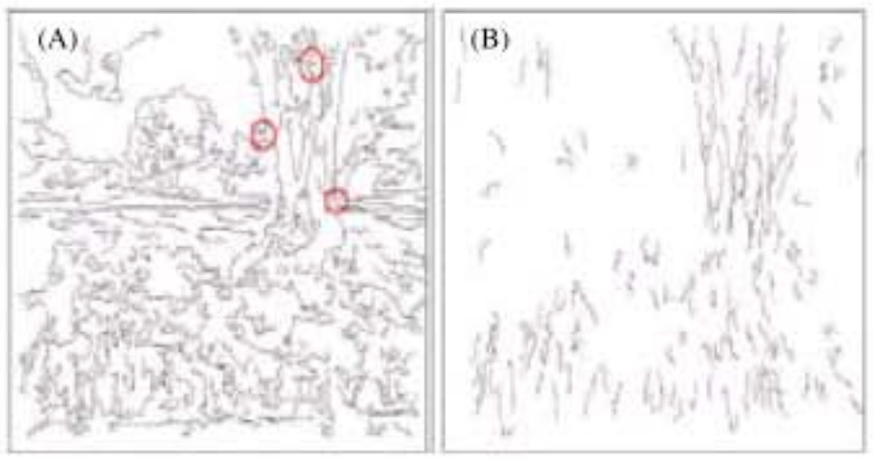

Fig. 11: (A) Traditional Canny compared with (B) Vertical Canny by Asmar *et al*. (2006)

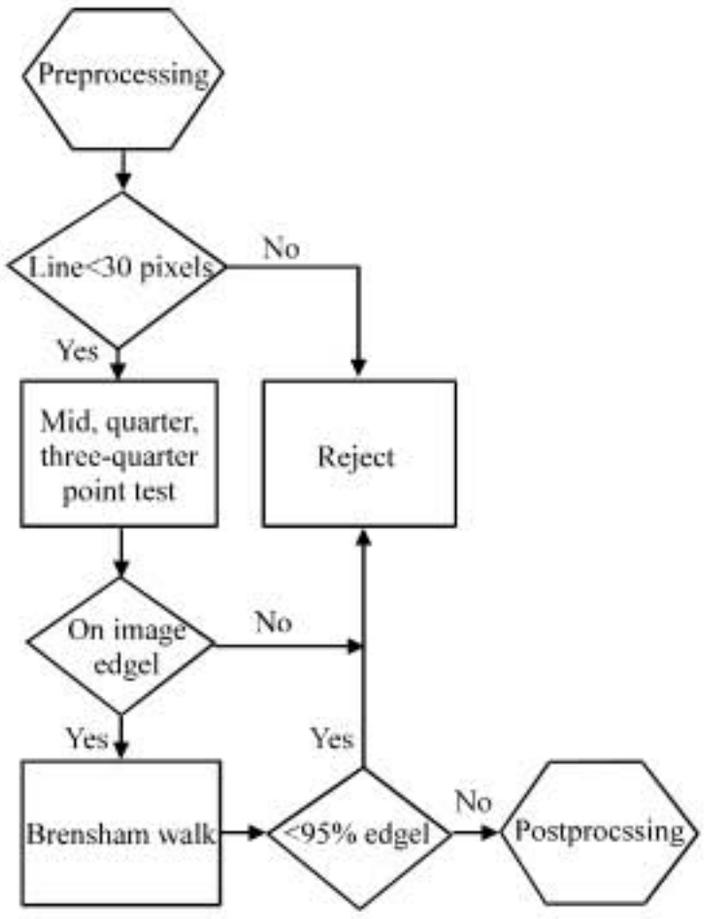

Fig. 12: Flow chart of Smith *et al*. (2006) implementation

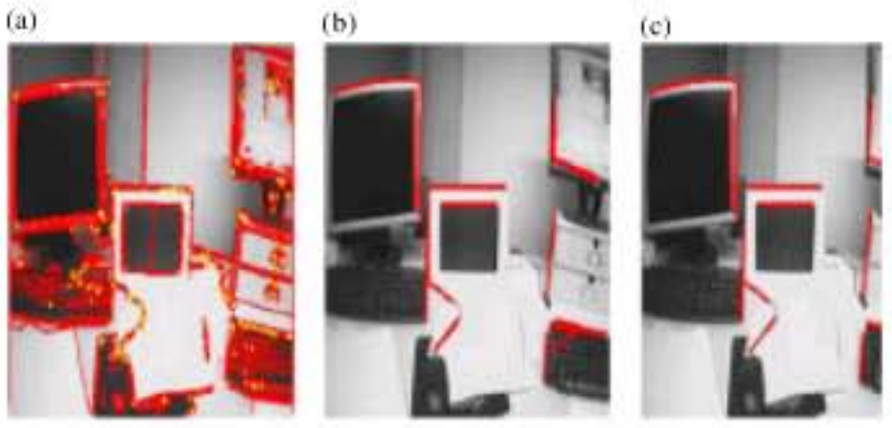

Fig. 13: Smith *et al*. (2006) result of (a) preprocessing, (b) mid processing and (c) post processing

VISION SLAM WITH SYSTEM ON CHIP APPROACH

System on Chip (SoC) is an approach that integrates multiple systems on a single silicon chip. Two most common term related to SoC are Application Specific Integrated Circuit (ASIC) and Field programmable gate array. Researchers and developers have long turned to special purpose hardware to accelerate image processing computation (Fowers *et al*., 2007). The computational power needed to satisfy actual embedded systems like video or image processing have accelerated the emergence of Systems On Programmable Chip (SoPC) (Chati *et al*., 2007). Besides light weight, low cost and low power solution, hardware implementation that makes use of parallel and pipelining processing provides a solution to complex calculation. This onboard solution also is desired to alleviate latency problems such occurred in video transmission process.

A research by Tippetts *et al*. (2007) of Brigham Young University has come out with an FPGA onboard solution. They exploit Harris Feature detector to calculate feature strengths and matches strengths with an x and y value representing the pixel location in the image. The original algorithm is scaled down to ensure that no overflow occurred. With Xilinx FX20 FPGA, the feature detector and priority queue detect and correspond the 20 strongest features at 30 fps. Studies by Chieh-Lun and Li-Chen (2007) also make use of Harris corner detector. Harris approach is identified as an image processing algorithm with a multilayer image process. Multi-layer image processing means that the final result image will takes more than once to process with different range of adjacent pixels or do the convolution with different kernel masks hierarchically. In CPU-based system, this multi-layer processing takes a lot of processing time and waste resources. Chieh-Lun and Li-Chen (2007) design a parallel computation structure by taking the advantage of FPGA pipeline and parallelism. The architecture of the design is show in Fig. 14.

An early FPGA implementation on SIFT algorithm was done by Se *et al*. (2004) for their planetary exploration rovers. FPGA is used since it able to speed the intensive image processing process and help to offload the processor. By comparison from a 640x480 image, SIFT feature extraction take 600 msec on a Pentium III 700 MHz processor while the FPGA (Xilinx Virtex II) is able to process within 60 msec and leaving the processor available for other task. Se *et al*. (2007) also compared their result with their earlier Tyzx stereo camera of 500x450 resolutions. The pure software systems runs at 2 Hz but improved to 4 Hz with Tyzx dense stereo card. With both the Tyzx dense stereo card and the SIFT FPGA, the system runs at 7 Hz. Following that, Barfoot (2005) use Se's SIFT feature extraction in their 3D motion estimation for SLAM. In their comparison, FPGA implementation is 6.6 times faster than the software implementation on Pentium-M to identify up to 2000 SIFT feature on 1024x768 image pixels. More recent study was done by Chati *et al*. (2007) where, hardware/software co-design

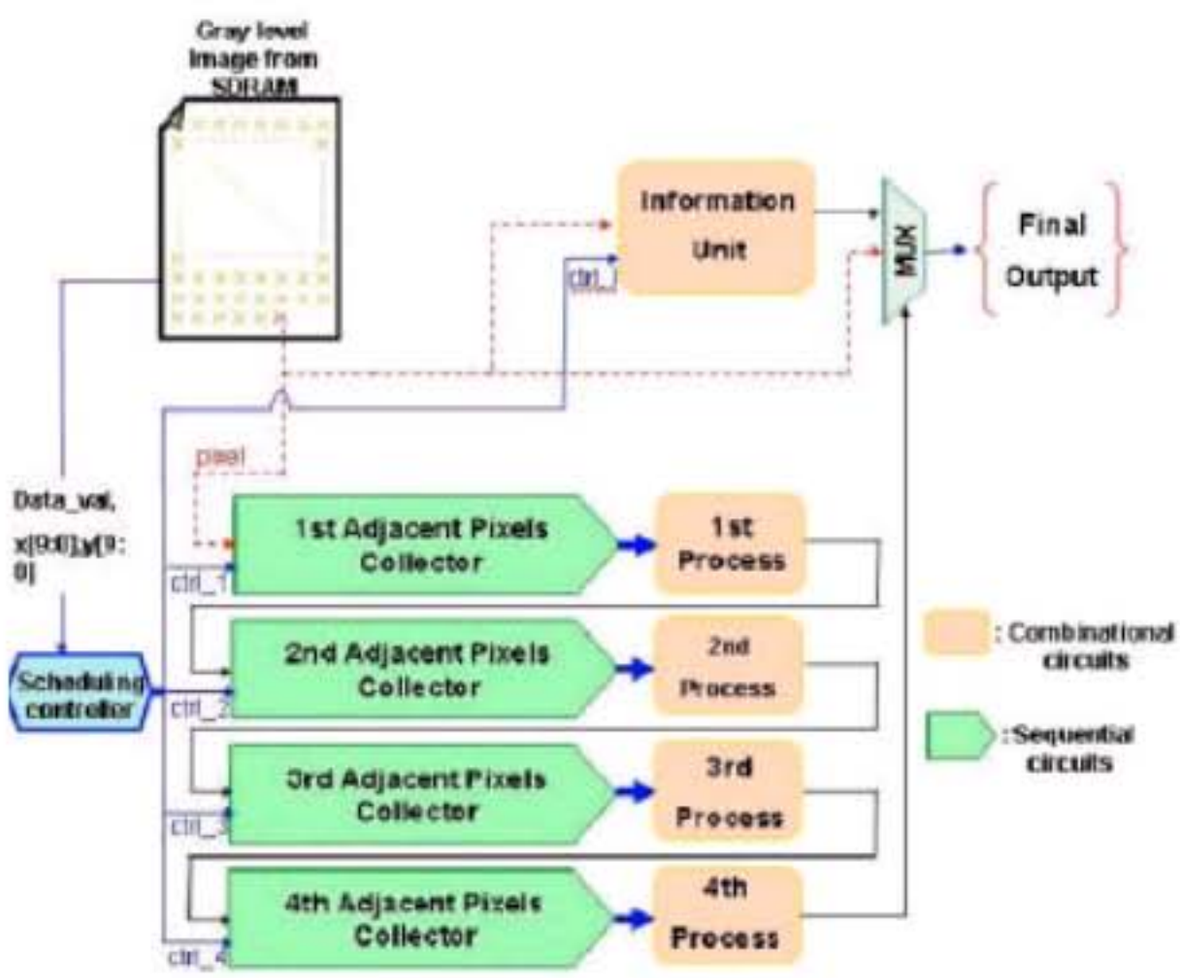

Fig. 14: The architecture of visual pipeline to realize the multi-layer image processing in real-time by Lu Chieh-Lun and Li-Chen (2007)

Table 4: Web-Link of Demo for feature detection

Approach	Web-Link of On-Line Source Code/Demo	File type	Contributor
Harris	http://www.mathworks.com/matlabcentral/ fileexchange/loadFile.do?objectId=9272	Matlab	Ali Ganoun
Harris, SUSAN, Harris-Laplace, Laplacian of Gaussian (LoG), Gilles	http://www.mathworks.com/matlabcentral/fileexchange/ loadFile.do?objectId=17894 and objectType=File	Matlab	Vincent Garcia, Universite De Nice
SIFT	http://www.cs.ubc.ca/~lowe/keypoints/	Matlab/C	David Lowe
	http://vision.ucla.edu/~vedaldi/code/sift/sift.html	Matlab/C	Andrea Vedaldi, University of California, LA
	http://robots.stanford.edu/cs223b04/MatlabSIFT.zip	Matlab	Stanford University
SURF	http://www.vision.ee.ethz.ch/~surf/download.html	Precompiled library/C++	Computer Vision Labora- tory, ETH Zurich
	http://www.csc.kth.se/utbildning/kth/kurser/DD2427/bik08	Matlab exercise	Stockholm University
FAST	http://svr-www.eng.cam.ac.uk/~er258/work/fast.html	Matlab/C	Edward Rosten
Phase congruency and spatial detection	http://www.csse.uwa.edu.au/~pk/research/matlabfns/	Matlab	The University of Western Australia

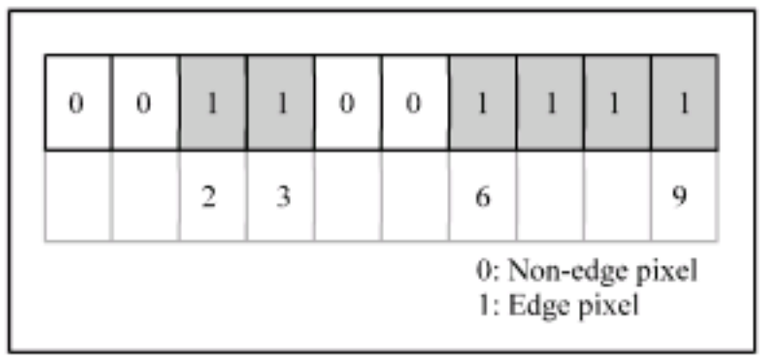

Fig. 15: Run length represented as (2,3) and (6,9)

of a SIFT key point detector on FPGA is presented. They claimed that convolution operation can be very efficiently implemented in hardware using sliding windows, where, multiplication can be done in parallel. This hardware/software partitioning and data streaming approach between modules is superior to processor approach which has to load, calculate and store each pixel value for each convolution. The hardware solution has only an initial delay and then can perform one pixel per clock.

Hariyama *et al*. (2008), they present their algorithm which consists of some tasks with high degree of column level parallelism in pre-processing, a window for stereo matching and post-processing. Even though the study describes FPGA implementation for vehicle detection, with some modification it can also be used in landmark detection. Sum of Absolute Difference (SAD) based matching and simplified 1D sobel operator is used in their implementation. One-dimensional sobel operator increases the degree of parallelism. To reduce memory accesses and computational amount, edge pixels are represented by a run-length representation as in Fig. 15.

CONCLUSIONS

This study has presented several key approach of feature detection with the current improvement done by several researchers. Feature detection is one of the main research areas for solving SLAM problems. As can be observed interest point detection is popular among researcher. However, the fusion of en edge and point detection approach has become significantly viable in the current research. To increase the robustness and real time implementation of feature detection technique, a system on chip approach is introduced in the SLAM community. The ability of the reprogrammable device to be marketed in time has made FPGA a favorable approach to be considered. The web link of source code and demo programs for feature detection approach is listed out in Table 4 (as accessed on 4th July 2008).

REFERENCES

Asmar, D.C., J.S. Zelek and S.M. Abdallah, 2006. Tree trunks as landmarks for outdoor vision slam. Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06), June 17-22, IEEE Computer Society, Washington, DC. USA., pp: 196-196.

Bar-Shalom, Y., X.R. Li and T. Kirubarajan, 2001. Estimation with Applications to Tracking and Navigation. 1st Edn., John Wiley and Sons, New York, ISBN: 0-471-41655-X.

Barfoot, T.D., 2005. Online visual motion estimation using fastslam with sift features. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005), August 2-6, pp: 579-585.

Bay, H., A. Essa, T. Tuytelaarsb and L.V. Goola, 2008. Speeded-Up Robust Features (SURF). Comput. Vision Image Understand., 110: 346-359.

Beis, J.S. and D.G. Lowe, 1997. Shape indexing using approximate nearest-neighbor search in high dimensional spaces. Conference on Computer Vision and Pattern Recognition, June 17-19, Puerto Rico, pp: 1000-1006.

Canny, J., 1986. A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell., 8: 679-698.

Castle, R.O., D.J. Gawley, G. Klein and D.W. Murray, 2007. Towards simultaneous recognition, localization and mapping for hand-held and wearable cameras. IEEE International Conference on Robotics and Automation, April 10-14, Rome, Italy, pp: 4102-4107.

Chati, H.D., F. Muhlbauer, T. Braun, C. Bobda and K. Berns, 2007. Hardware/Software co-design of a key point detector on FPGA. International Symposium on Field-Programmable Custom Computing Machines, April 23-25, Napa, CA, pp: 355-356.

Chatila, R. and J.P. Laumond, 1985. Position referencing and consistent world modeling for mobile robots. Proceedings of the IEEE International Conference on Robotics and Automation, Mar. 25-28, IEEE Xplore London, pp: 138-143.

Chieh-Lun, L. and F. Li-Chen, 2007. Hardware architecture to realize multi-layer image processing in real-time. The 33rd Annual Conference of the IEEE Industrial Electronics Society (IECON), Nov. 5-8, Taipei, Taiwan, pp: 2478-2483.

Chum, O. and J. Matas, 2005. Matching with PROSAC-progressive sample consensus. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, Los Alamitos, USA., pp: 220-226.

Clemente, L.A., A.J. Davison, I.D. Reid, J. Neira and J.D. Tardos, 2007. Mapping Large Loops with a Single Hand-Held Camera Robotics: Science and Systems (RSS). 1st Edn., The MIT Press, Atlanta, ISBN: 978-0-262-52484-1.

David, G.L., 1999. Object recognition from local scale-invariant features. Proceeding of the International Conference on Computer Vision, IEEE Computer Society Washington, DC. USA., pp: 1150-1157.

David, G.L., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision, 60: 91-110.

Davison, J. and D.W. Murray, 2002. Simultaneous localization and map-building using active vision. IEEE Trans. Pattern Anal. Mach. Intell., 24: 865-880.

Davison, J., 2003. Real-time simultaneous localisation and mapping with a single camera. Proceedings of the 9th IEEE International Conference on Computer Vision (ICCV'03), Oct. 13-16, IEEE Xplore, London, pp: 1403-1410.

Deriche, R., 1987. Using canny's criteria to derive an optimal edge detector recursively implemented. Int. J. Comput., 1: 167-187.

Dissanayake, M.W.M.G., P. Newman, S. Clark, H.F. Durrant-Whyte and M. Csorba, 2001. A solution to the simultaneous localization and map building (SLAM) problem. IEEE Trans. Robotics Automation, 17: 229-241.

Durrant-Whyte, H. and T. Bailey, 2006. Simultaneous localisation and mapping (SLAM): Part I the essential algorithms. Robotics Automa. Magazine, 13: 99-108.

Eade, E. and T.W. Drummond, 2006. Edge landmarks in monocular slam. Proceeding BMVC'06, Sept. 2006, Edinburgh, Scotland, UK., pp: 1-10.

Escolano, F., B. Bonev, P. Suau, W. Aguilar, Y. Frauel, J.M. Saez and M. Cazorla, 2007. Contextual visual localization: Cascaded submap classification, optimized saliency detection and fast view matching. International Conference on Intelligent Robots and Systems IEEE/RSJ, Oct. 29-Nov. 2, San Diego, CA, USA., pp: 1715-1722.

Fowers, S.G., L. Dah-Jye, B.J. Tippetts, K.D. Lillywhite, A.W. Dennis and J.K. Archibald, 2007. Vision aided stabilization and the development of a quad-rotor micro UAV. International Symposium on Computational Intelligence in Robotics and Automation CIRA, June 20-23, Jacksonville, FI, pp: 143-148.

Frintrop, S., 2006. VOCUS: A Visual Attention System for Object Detection and Goal-Directed Search. LNCS 3899. Springer, Heidelberg, ISBN: 978-3-540-32759-2.

Frintrop, S., P. Jensfelt and H. Christensen, 2007. Simultaneous Robot Localization and Mapping Based on a Visual Attention System. In: WAPCV LNAI 4840, Paletta, L. and E. Rome (Eds.). Springer-Verlag, Berlin Heidelberg, pp: 417-430.

Fu, S., H.Y. Liu, L.F. Gao and Y.X. Gai, 2007. SLAM for mobile robots using laser range finder and monocular vision. Proceedings of the 14th International Conference on Mechatronics and Machine Vision in Practice, Dec. 4-6, Xiamen, pp: 91-96.

Haralick, R.M., C.N. Lee, K. Ottenberg and M. Nolle, 1994. Review and analysis of solutions of the three point perspective problem. Int. J. Comput. Vision, 13: 91-110.

Harati, A., S. Gachter and R. Siegwart, 2007. Fast range image segmentation for indoor 3D-SLAM. Proceeding of 6th IFAC Symposium on Intelligent Autonomous Vehicle (IAV), Sept. 3-5, Toulouse, France, pp: 1-6.

Hariyama, M., K. Yamashita and M. Kameyama, 2008. FPGA implementation of a vehicle detection algorithm using three-dimensional information. IEEE International Symposium on Parallel and Distributed Processing (IPDPS), April 14-18, IEEE Xplore, London, pp: 1-5.

Harris, C. and M. Stephens, 1988. A combined corner and edge detector. Proceedings of the 4th Alvey Vision Conference, (AVC'88), Manchester, UK., pp: 147-151.

Hygounenc, E., I.K. Jung, P. Soueres and S. Lacroix, 2004. The autonomous blimp project of laas-cnrs: Achievements in flight control and terrain mapping. Int. J. Robotics Res., 23: 473-511.

Idna, M.Y. and E.M. Tamil, 2007. Parking information system using GPS and shortest path algorithm. Proceedings of the SCORED 2007, May 14-15, Universiti Tenaga Nasional, Malaysia, pp: 1-7.

Idna, M.Y., E.M. Tamil and Z. Razak, 2008. Reducing emergency vehicle response time using GPS technology and shortest path algorithm. Proceedings of Internet Convergence Conference (ICC 2007), Mar. 11-13, 2008, Kuala Lumpur. Malaysia.

Ke, Y. and R. Sukthankar, 2004. PCA-SIFT: A more distinctive representation for local image descriptors. Comput. Vision Pattern Recognition, 2: 506-513.

Lemaire, T. and S. Lacroix, 2007. Monocular-vision based SLAM using line segments. Proceedings of the IEEE International Conference on Robotics and Automation, April 10-14, IEEE Xplore, London, pp: 2791-2796.

Lemaire, T., C. Berger, I.K. Jung and S. Lacroix, 2007. Vision-Based SLAM: stereo and monocular approaches. Int. J. Comput. Vision, 74: 343-364.

Mikolajczyk, K. and C. Schmid, 2001. Indexing based on scale invariant interest points. Proceedings of the 8th International Conference on Computer Vision, 2001 Vancouver, Canada, pp: 525-531.

Mikolajczyk, K. and C. Schmid, 2002. An affine invariant interest point detector. Proceedings of the 7th European Conference on Computer Vision-Part I, LNCS 2350, 2002, Springer-Verlag, pp: 128-142.

Mikolajczyk, K. and C. Schmid, 2004. Scale and affine invariant interest point detectors. Int. J. Comput. Vision, 60: 63-86.

Mikolajczyk, K. and C. Schmid, 2005. A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell., 27: 1615-1630.

Mirisola, L.G.B., J. Dias and A.T. De Almeida, 2007. Trajectory recovery and 3D mapping from rotation-compensated imagery for an airship. Proceedings of the IEEE/RSJ International Conference on Intel. Robots and System, Oct. 29-Nov. 2, San Diego, CA, USA., pp: 1908-1913.

Mountney, P., D. Stoyanov, A. Davison and G.Z. Yang, 2006. Simultaneous stereoscope localization and soft-tissue mapping for minimal invasive surgery. Proceedings of the 9th International Conference, Copenhagen, Denmark, LNCS 4190, Oct. 1-6, Springer Berlin/Heidelberg, pp: 347-354.

Nieto, J., T. Bailey and E. Nebot, 2007. Recursive scan-matching slam. Robotics Autonomous Syst., 55: 39-49.

Park, Y., S. Jeong, I.H. Suh and B.U. Choi, 2007. Map-Building and Localization by Three-Dimensional Local Features for Ubiquitous Service Robot. In: ICUCT 2006, LNCS., 4412, Stajano, F. *et al.* (Eds.). Springer-Verlag, Berlin/Heidelberg, pp: 69-79.

Quinlan, J.R., 1986. Induction of decision trees. Machine Learn., 1: 81-106.

Rosten, E. and T. Drummond, 2005. Fusing points and lines for high performance tracking. Int. Conf. Comput. Vision, 2: 1508-1515.

Rosten, E. and T. Drummond, 2006. Machine learning for high-speed corner detectio. Proceedings of the 9th European Conference on Computer Vision, Graz, Austria, LNCS 3951, May 7-13, Springer Berlin/Heidelberg, pp: 430-443.

Royer, E., M. Lhuillier, M. Dhome and J.M. Lavest, 2007. Monocular vision for mobile robot localization and autonomous navigation. Int. J. Comput., 74: 237-260.

Schleicher, D., L.M. Bergasa, R. Barea, E. Lopez, M. Ocaa, J. Nuevo and P. Fernandez, 2007. Real-time stereo visual slam in large-scale environments based on sift fingerprints. IEEE International Symposium on Intelligent Signal Processing, WISP 2007, Oct. 3-5, IEEE Xplore, London, pp: 1-6.

Schleicher, D., L.M. Bergasa, R. Barea, E. Lopez, M. Ocana and J. Nuevo, 2007. Real-time wide-angle stereo visual slam on large environments using sift features correction. Proceedings of the IEEE/RSJ International Conference on International Robots and System, Oct. 29-Nov. 2, San Diego, CA, USA., pp: 3878-3883.

Schmid, C., R. Mohr and C. Bauckhage, 2000. Evaluation of interest point detectors. Int. J. Comput., 37: 151-172.

Se, S., H. Ng, P. Jasiobedzki and T. Moyung, 2004. Vision based modeling and localization for planetary exploration rovers. 55th International Astronautical Congress Vancouver, Canada, Oct. 4-8, Curran Associates Inc., pp: 1-11.

Shaw, D. and N. Barnes, 2006. Perspective rectangle detection. 9th European Conference on Computer Vision (ECCV2006), Graz, Austria, May 7-13, 2006.

Shen, Y. and J. Liu, 2007. Vision based SLAM for robot navigation with single camera. International Scientific and Technology Exhibition Congress Mechatronics and Robotics (M and R-2007), Oct. 2-5, St. Petersburg, pp: 1-10.

Smith, P., I. Reid and A. Davison, 2006. Real-time monocular slam with straight lines. Br. Mach. Vision Conf., 1: 17-26.

Smith, R. and P. Cheesman, 1987. On the representation and estimation of spatial uncertainty. Int. J. Robotics Res., 5: 56-68.

Smith, R., M. Self and P. Cheeseman, 1988. A stochastic map for uncertain spatial relationships. Robotics Research: The 4th International Symposium, 1988, The MIT Press, Cambridge, MA, USA., pp: 467-474.

Smith, R., M. Self and P. Cheeseman, 1990. Estimating Uncertain Spatial Relationships in Robotics. In: Autonomous Robot Vehicles, Cox, I.J. and G.T. Wilfon (Eds.). Springer-Verlag, New York, ISBN:0-387-97240-4, pp: 167-193.

Steder, B., G. Grisetti, S. Grzonka, C. Stachniss, A. Rottmann and W. Burgard, 2007. Learning maps in 3D using attitude and noisy vision sensors. Proceedings of the International Conference on Intelligent Robots and Systems IEEE/RSJ, Oct. 29-Nov. 2, San Diego, CA, USA., pp: 644-649.

Strasdat, H., C. Stachniss, M. Bennewitz and W. Burgard, 2007. Visual Bearing-Only Simultaneous Localization and Mapping with Improved Feature Matching Autonome Mobile Systeme. 1st Edn., Springer, USA., ISBN: 3540747648.

Taylor, T., 2007. Applying high-level understanding to visual localisation for mapping springer berlin/heidelberg. Autonomous Robots Agents, 76: 35-42.

Tippetts, B., S. Fowers, K. Lillywhite, L. Dah-Jye and J. Archibald, 2007. FPGA implementation of a feature detection and tracking algorithm for real-time applications. Proceedings of the 3rd International Symposium, ISVC 2007, Lake Tahoe, NV, USA, LNCS 4841, Nov. 26-28, Springer Berlin/Heidelberg, pp: 682-691.

Tomasi, C. and T. Kanade, 1991. Shape and motion from image streams: A factorization method-part 3. Detection and Tracking of Point Features Technical Report CMU-CS-91-132.

Tuytelaars, T. and K. Mikolajczyk, 2008. Local invariant feature detectors: A survey. Foundat. Trends Comput. Graphics Vision, 3: 177-280.

Wijk, O., 2001. Triangulation Based Fusion of Sonar Data with Application in Mobile Robot Mapping and Localization. 1st Edn., Stockholm: KTH, Signals, Sensors and System, UK., ISBN: 91-7283-054-9.

Williams, B., P. Smith and I. Reid, 2007. Automatic relocalisation for a single-camera simultaneous localisation and mapping system. International Conference on Robotics and Automation, April 10-14, Rome, Italy, pp: 2784-2790.

Yong-Ju, L. and S. Jae-Bok, 2007. Autonomous selection, registration and recognition of objects for visual slam in indoor environments. COEX, International Conference on Control, Automation and System, Oct. 17-20, Seoul, Korea, pp: 668-673.

Zunino, G., 2006. Simultaneous Localization and Mapping for Navigation in Realistic Environments. 1st Edn., Publisher KTH, New York, ISBN: 91-7283-246-0.