Autonomous Flight in Unstructured Environment using Monocular Image

Saurav Kumar[1], Ashutosh Saxena[2]
Department of Computer Science, Cornell University
[1]sk2357@cornell.edu [2]asaxena@cs.cornell.edu

I. Introduction

Over the past decade various research work have successfully allowed unmanned aerial vehicle to autonomously explore and simultaneously map different kinds of environment ranging from outdoor to indoor environment, from learning trajectory to [1,2,3,4]. There have been numerous efforts in the past to fly unmanned aerial vehicle using a single camera. While many works have demonstrated autonomous maneuvers but the work is limited to environment with specific features for e.g. flying in corridor where vanishing point is a strong cue to emphasize on heading of the aerial platform. SLAM also becomes one of the major issues thought it has been proven very successfully for the ground vehicle primarily because of the fact that MAV generally have low weight carrying capacity. It can't carry high aperture camera or popular SICK LRF or Industrial Graded IMU [5,6,7].

Work mentioned in [6] describes an approach that uses optical flow information to estimate the objects in the scene, which are near to the aerial vehicle platform. However the optical flow is highly affected by rotation of the camera scene thus they used a high precise IMU to take care of the rotational image. Another bottleneck is experienced with the computation power of the system. Wireless transmission of images to a base station for feature extraction and matching process is one of solution since we aren't considering a very fast motion it works for slow moving and stable ground vehicle but its stills an issue for dynamically unstable helicopters. Transferring image over wireless becomes a bottleneck since there are transmission delays and also since we get rid of high frequency component of image by applying a lossy compression before transmitting image, it becomes difficult to feature detector algorithm to work on. Path planning for MAV becomes another issue. When inaccuracy resulting from not very accurate sensing system contains large uncertainty, the MAV can't just simply stop and perform more sensing. Instead the MAV will probably be unable to estimate its own velocity accurately, as result might pick up speed or oscillate, degrading the measurement further. Since MAV flies totally in a 3D environment so the visible 2D slice of the environment can drastically with height and attitude as obstacle might suddenly appear or disappear.

II. Motivation and Learning Approach

Main motive of the project is to be able to fly an indoor helicopter figure 1(a) with input only from one monocular front camera and use sonar for height detection. So we need to estimate depth information for different regions for the monocular image so that a controller can predict

Figure 1a) Aerial Vehicle used for experiment **b) Visualization of depth on single image**

Free space and enable quad rotor to fly autonomously without hitting into an object which is closer to the camera. The visualization of the work is shown in figure 1(b). We investigated into different features that could give us information about depth of an object into the scene. Some of the monocular cues that provides us with depth information are

1. Texture gradient: - Suppose you are standing on a gravel road. The gravel near you can be clearly seen in terms of shape, size and color. As your vision shifts towards the distant road the texture cannot be clearly differentiated. In the figure 2 below, a clarity image is obtained from gray image on the left.

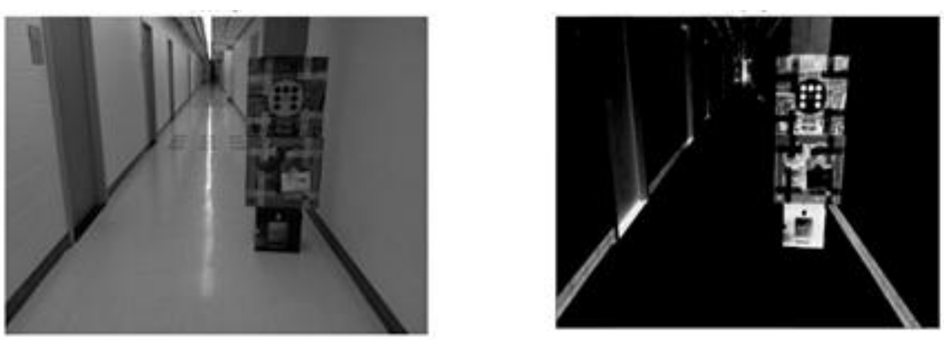

Figure 2 : Gray Image and Clarity Image

2. Motion parallax - When an observer moves, the apparent relative motion of several stationary objects against a background gives hints about their relative distance. If information about the direction and velocity of movement is known, motion parallax can provide absolute depth information. Parallax is defined as apparent displacement, or difference in the apparent position, of an object, caused by actual change (or

difference) of position of the point of observation; spec. the angular amount of such displacement or difference of position, being the angle contained between the two straight lines drawn to the object from the two different points of view, and constituting a measure of the distance of the object [12]. This effect can be seen clearly when driving in a car. Nearby things pass quickly, while far off objects appear stationary. Some animals that lack binocular vision due to wide placement of the eyes employ parallax more explicitly than humans for depth cueing (e.g. some types of birds, which bob their heads to achieve motion parallax, and squirrels, which move in lines orthogonal to an object of interest to do the same). It could be further understood that when an object is viewed from two widely separated positions, the object in front appears to move faster with respect to more distant objects. This is true for observations on earth (indoor and outdoor environments) and also for observations of stars & planets. Figure 3 demonstrate the concept of parallax in the actual test scenario where left pair of image is taken from a height greater than the height of the camera in the right side of the image. If you will closely into the image you will find that the object which was closer to the camera has moved much in the scene between two pair of images than the object that was in the foreground. This concept serves as the basic of our machine learning information.

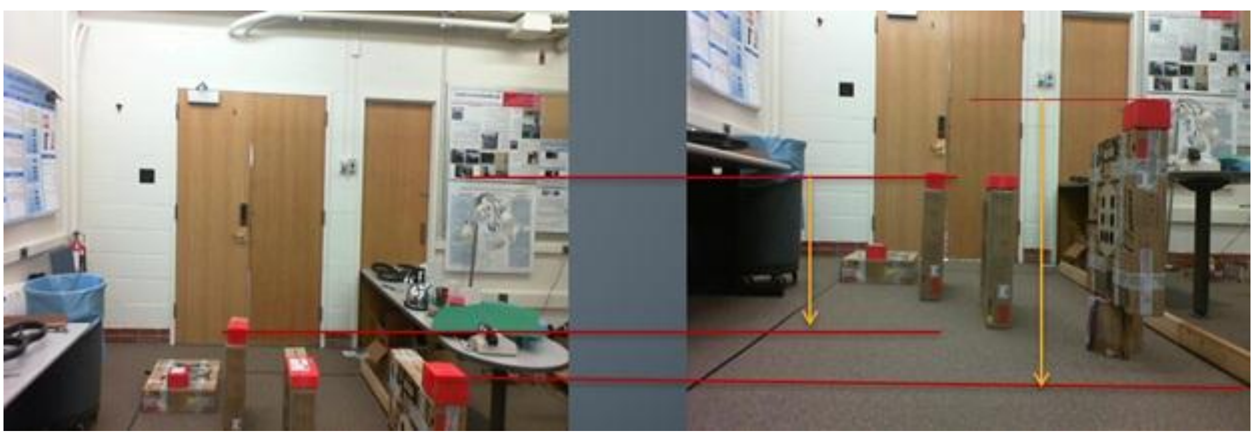

Figure 3: Demonstration of Parallax in the scene.

3. Depth from motion - One form of depth from motion, kinetic depth perception, is determined by dynamically changing object size. As objects in motion become smaller, they appear to recede into the distance or move farther away; objects in motion that appear to be getting larger seem to be coming closer. Using kinetic depth perception enables the brain to calculate time to crash distance (aka time to collision or time to contact - TTC) at a particular velocity. When driving, we are constantly judging the dynamically changing headway (TTC) by kinetic depth perception.

Estimation of motion becomes quite challenging to estimate since it's difficult to estimate the correct correspondence between the pixels in the scene as there exists a lot ambiguity in the scene since for e.g. a pixel on the wall on left image can be almost any other wall pixel computationally. This problem is aggravated by the fact that there exist very less SIFT matches in between the images since SIFT doesn't detect pixels which has ambiguity as could be shown in figure 4 below.

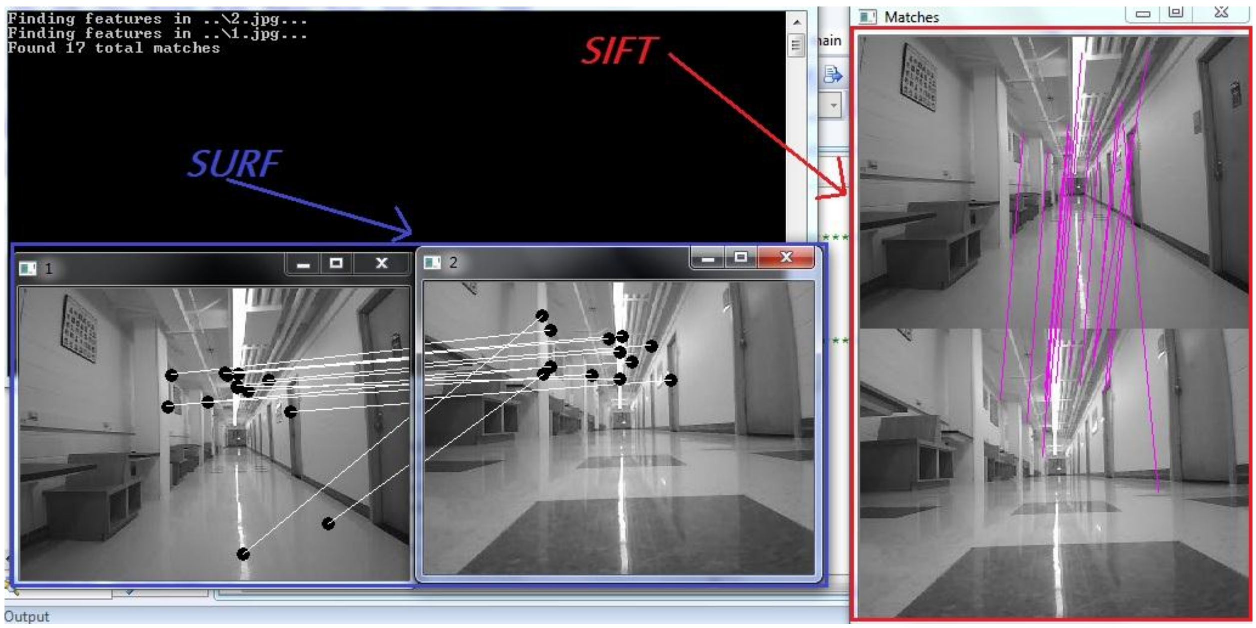

Figure 4: Fewer Number of SIFT/SURF matches.

To solve this problem, we took help from Structure of Motion methodology. We tried to estimate a fundamental matrix using the known SIFT matches to predict the motion of camera in the scene. The figure below shows the SIFT matching approach.

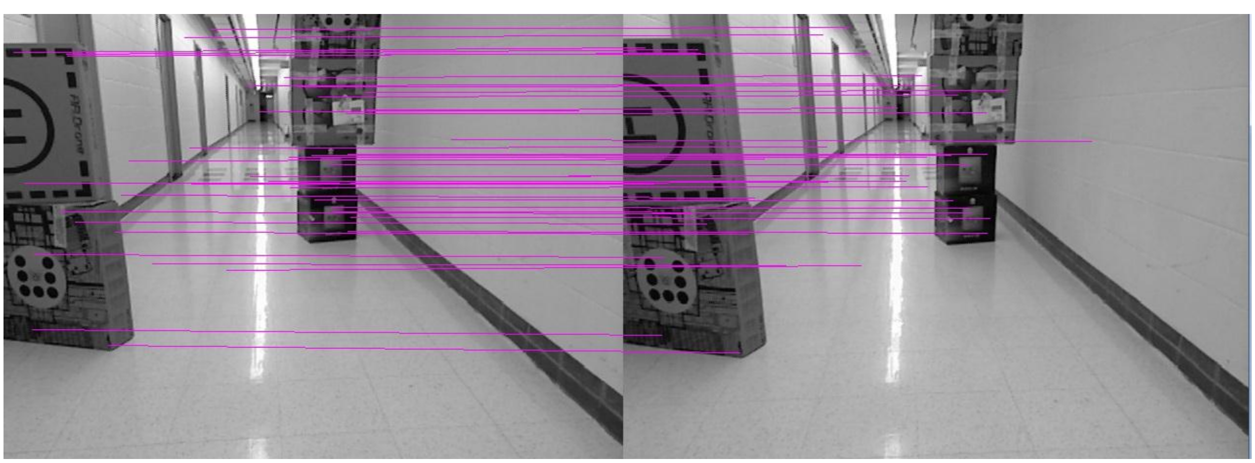

Figure 5: SIFT matches is used to estimate Fundamental Matrix

We calculate the fundamental matrix using 8 Point Algorithm. This matrix maps a pixel in one image to lines in the other image as could be seen from figure 6 where the point on the left image as shown by red circle maps to green lines in the right image.

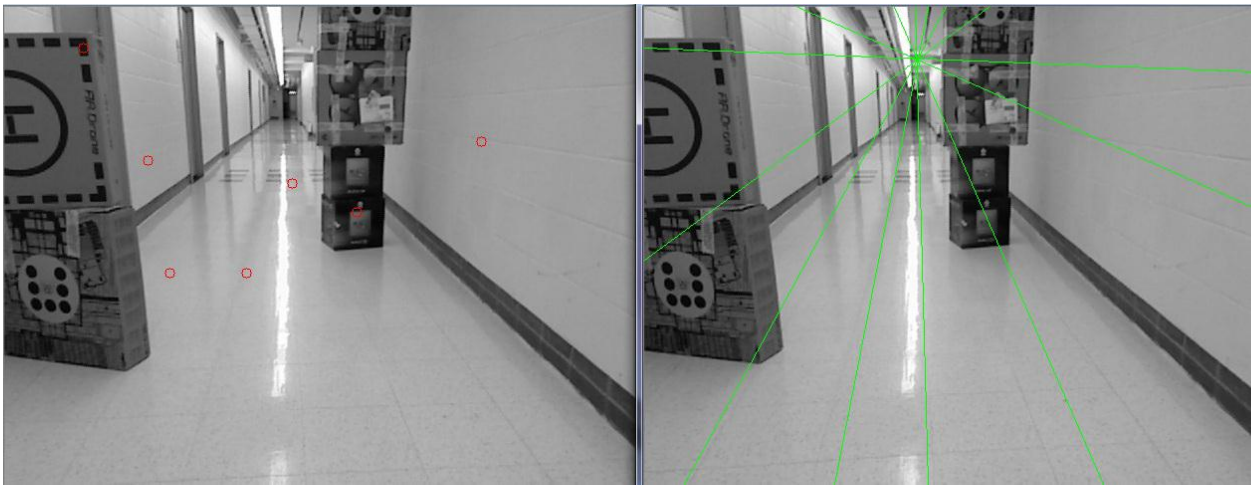

Figure 6: Point – Line Correspondence in subsequent images

We then use a block matching approach to find the actual pixel and estimate dense optical flow for the scene and that gives us an estimate of movement of different pixels in the image and that will provide us with vital cue for depth perception.

III. Dataset Generation and Supervised Learning

To estimate depth from monocular image, we created a dataset of over 4000 gray images and its corresponding depth image using Kinect. Some of the sample images obtained from the experiment is shown in figure 8. For training of our algorithm, for each gray image we compute different feature vector like Clarity image , Contrast , Energy , Edge , Surface image etc as shown in figure 7.

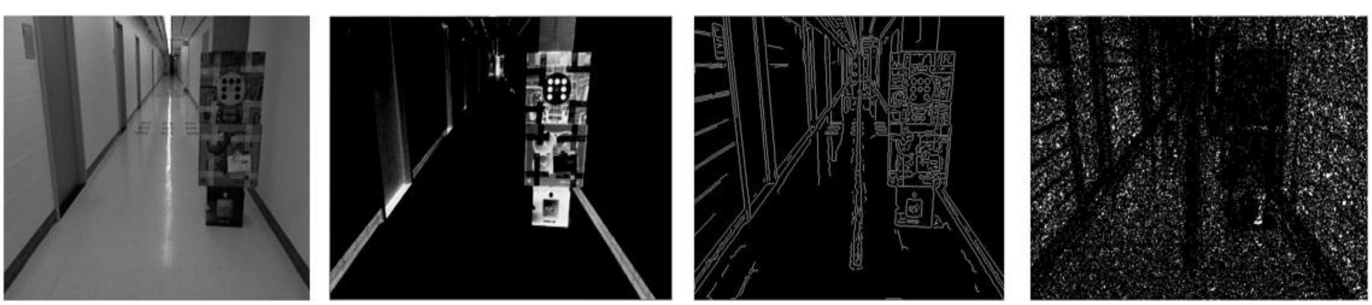

Figure 7: Gray Image, Clarity Image, Edge Image, Surface image

Figure 8: Gray and Depth Image captured using Kinect

However we had to improve the depth map as shown above to a more sampled version. To determine depth for a patch in an image, we assigned the lowest depth associated over the entire pixel in that patch. The refined depth map thus obtained is shown in figure 9 for one of a sample depth map. This refined depth map is used for the training our learning algorithm to predict depth in the scene.

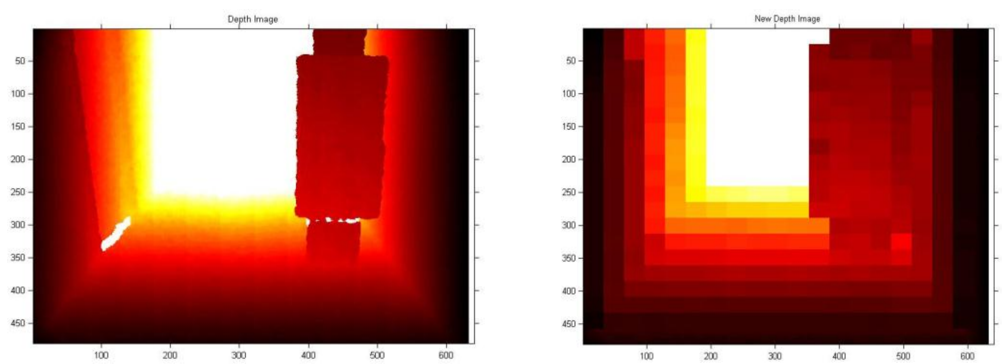

Figure 9: Original Depth Map and Refined Depth Map

IV. **Result and Future work**

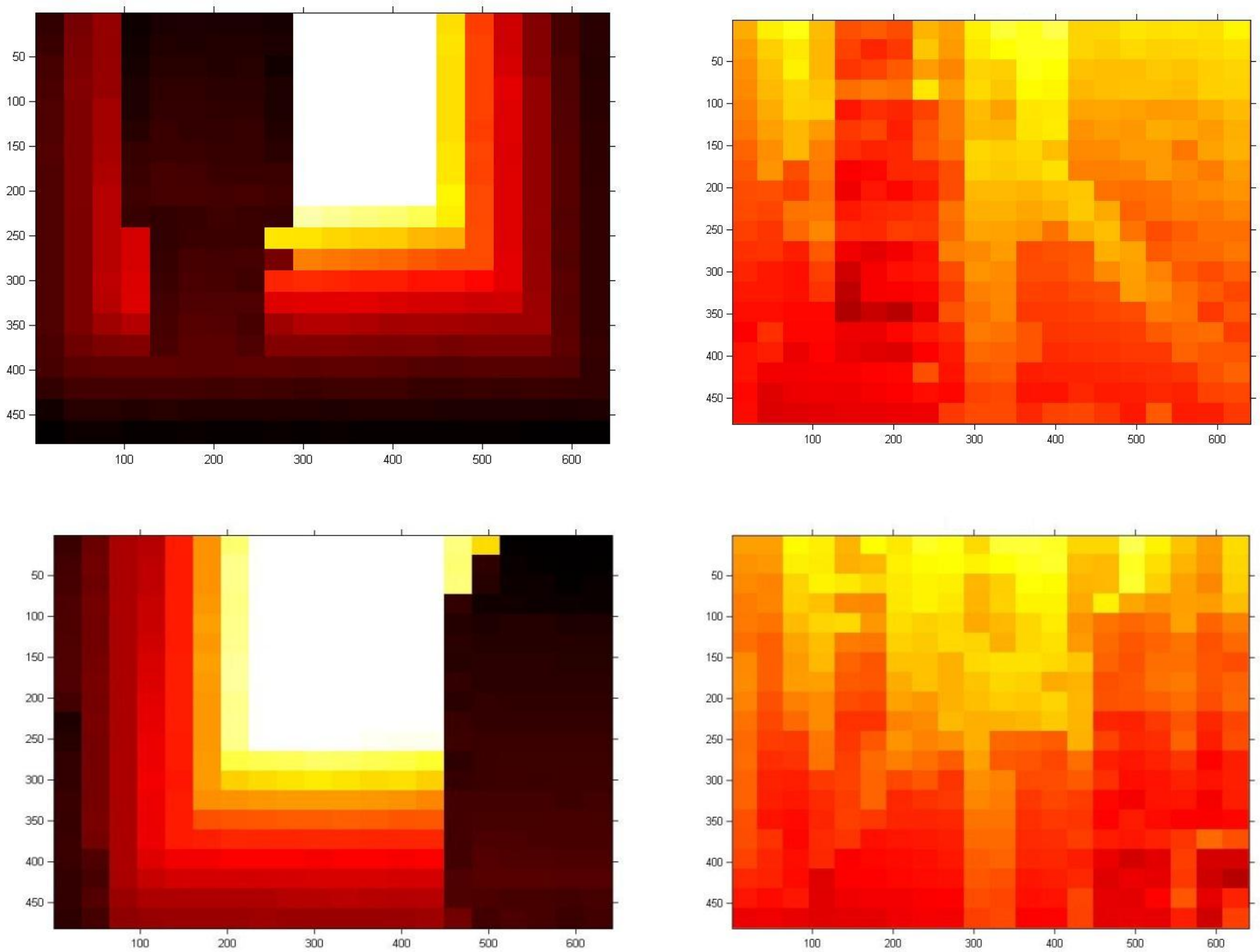

Figure 10: Original Depth Map and Predicted Depth Map

The Initial result over the predicted depth map is obtained using Linear Regression with 23 Features. We built separate dataset for training and test images. The result is obtained over test image. Detection of obstacle in scene and floor could be visualized easily from the depth map. The free space region denoted by yellow color in the image is also predicted in correct way.

Few top most rows could be interesting rows to predict about the area containing free space and decide about the heading direction for the Quadrotor. A visualization of the future work is shown in figure 11 where the top 5 rows are used to create a grid representation of environment with obstacle denoted as red and free space with white. Heading direction could be determined using a simple controller relying on potential field method like VFH.

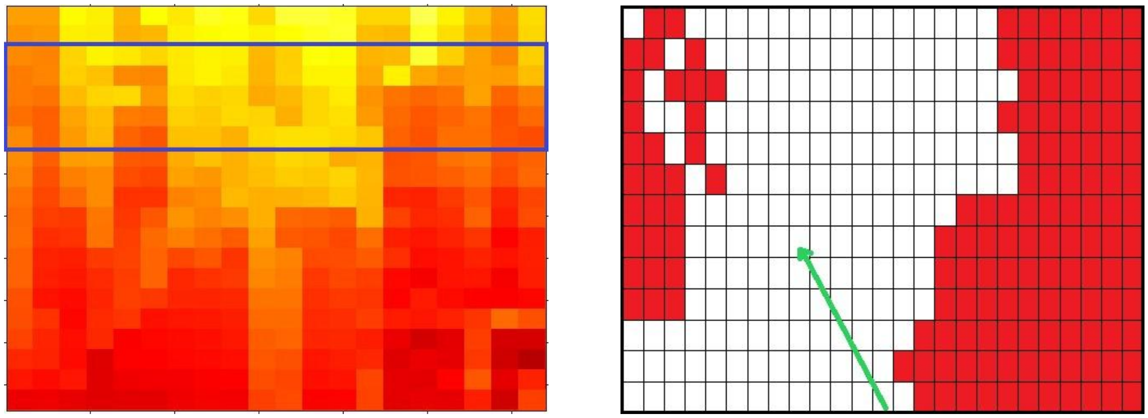

Figure 11: Predicted Depth Map and Grid Formation Visualization

Currently the result is shown using considering cues like contrast gradient, edge, surface map etc. However in the immediate future we will implement motion cue like, parallax, kinetic energy, momentum of different patch to further improve the result. In the future we will also try to learn the mapping function with different weights for different features and also try to come with learning algorithm that could help us fuse information looking at different scale of the image.

On a final note, in this study, we addressed the problem of obstacle avoidance using vision based methods, where the navigation of a miniature Quadrotor helicopter within an indoor corridor was the main task. Relying on motion based depth map could prove out an interesting area to look into. I believe that depth-map based collision avoidance may evolve as a very powerful tool, since it might be applied in unknown environments with a high density of objects, and therefore be ideal for use in indoor environments.

References:-

1. Myungsoo Jun, Stergios I. Roumeliotis, and Gaurav S. Sukhatme. "State Estimation of an Autonomous Helicopter using Kalman Filtering". In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 1346-1353, Kyongju, Korea, 1999.
2. Srikanth Saripalli, Gaurav S. Sukhatme, and James F. Montgomery. "An Experimental Study of the Autonomous Helicopter Landing Problem". In Proceedings, International Symposium on Experimental Robotics (ISER), Sant'Angelo d'Ischia, Italy, 2002
3. Pieter Abbeel pabbeel, Adam Coates ,Andrew Y. Ng, "Autonomous Helicopter Aerobatics through Apprenticeship Learning" , International Journal of Robotics Research 2010
4. F. Ruffier, and N. Franceschini, "Visually guided micro-aerial vehicle: automatic take off, terrain following, landing and wind reaction," International Conference on Robotics and Automation (ICRA), 2004.
5. S. Bouabdallah and R. Siegwart. “Backstepping and sliding-mode techniques applied to an indoor micro quadrotor”, International Conference on Robotics and Automation (ICRA), 2005.

6. Simon Zingg, Davide Scaramuzza, Stephan Weiss, Roland Siegwart , "MAV Navigation through Indoor Corridors Using Optical Flow", International Conference on Robotics and Automation (ICRA), 2010.
7. Sai Prashanth Soundararaj, Arvind K. Sujeeth and Ashutosh Saxena,"Autonomous Indoor Helicopter Flight using a Single Onboard Camera”, International Conference on Robotics and Automation (ICRA), 2008.
8. Koray Çelik, Soon-Jo Chung , Matthew Clausman and Arun K. Somani , "Monocular vision SLAM for indoor aerial vehicles" , 2009 IEEE/RSJ international conference on Intelligent robots and systems
9. Ruijie He, Sam Prentice and Nicholas Roy , "Planning in Information Space for a Quadrotor Helicopter in a GPS-denied Environment", International Conference on Robotics and Automation (ICRA), 2008
10. P. Abbeel, A. Coates, M. Quigley, and A. Y. Ng, “An application of reinforcement learning to aerobatic helicopter flight.” Neural Information Processing Systems 2006.”
11. Abraham Bachrach, Ruijie He and Nicholas Roy , "Autonomous Flight in Unknown Indoor Environments" , IJMAV 2009
12. "Parallax", Oxford English Dictionary (Second Edition ed.). 1989.
13. Richard I. Hartley, "In Defense of the Eight-Point Algorithm", IEEE Transaction on Pattern Recognition and Machine Intelligence