A SIMPLE COMPUTER VISION BASED INDOOR POSITIONING SYSTEM FOR EDUCATIONAL MICRO AIR VEHICLES

Submitted: 5th May 2013; accepted: 2nd June 2013

Thomas Kittenberger, Andreas Ferner, Reinhard Scheikl

DOI: 10.14313/JAMRIS_1-2014/6

Abstract:

Computer vision is one of the main research fields for micro air vehicles. Therefore many teams around the globe are developing different camera systems which are directly mounted on a MAV or observe and track it from a fixed position. This paper presents a simple and cheap solution for a precise 40 gram onboard computer vision system based on a 6 gram Gumstix Overo Firestorm computer, a 23 gram infrared camera a custom adapter board including buck converter and connectors and three ground based high power infrared LED's. It has great potential for educational applications because no fixed and calibrated camera installation around the flight area is necessary so it can easily be deployed and removed. The use of various filter algorithms makes it very reliable and the 752x480 pixel images are processed with 12 fps. The achieved accuracy is about 2% and the range was tested up to 10 m.

Keywords: *computer vision, indoor positioning, Gumstix, infrared camera, high power LED, robotics education, MAV*

1. Introduction and Related work

In recent years small mobile ground based or aerial robots like line followers, sumo and soccer robots or quadrocopters have become a very popular means to teach technical topics with project based learning at universities. The Department of Embedded Systems at the University of Applied Sciences Technikum Wien founded its Micro Air Vehicle Team in 2007. The team developed different kinds of inertial, ultrasonic and GPS based navigation systems, flight control computers and aerial robots [1] [2] and won in 2010 and 2012 two third places at the RobotChallenge, the largest European robotics competition.

The type of contest we attended in 2012 was the new "Air Race". In the semi autonomous class the aircraft had to fly in 10 minutes as much figure eight patterns as possible around two poles with 5 m distance with computer controlled height. Only the flight path was controlled manually. Our team achieved 23 rounds with an attitude stabilized hovering mono rotor aircraft similar to Figure 1 and won the third place. For the next year's contest it was decided to participate at the autonomous Air Race where, for the same task, no human interaction is allowed [3]. So it was necessary to find an adequate solution for positioning our hovering aircrafts in an indoor space with a size of about 5x10x3 m which could be used at the contest but also in different large class rooms and labs at our university. After a first glance at different available technologies like ultra wide band radio, ultrasonic or vision based indoor positioning systems it was decided to use a simple vision based solution with some kind of easily detectable markers. This approach opens the way to a subsequent development for operation in natural environments without using explicit markers whereas the other solutions always rely on some kind of infrastructure in the flight area.

A simple and widely used approach to vision based indoor positioning is to use a downward oriented camera on the Micro Air Vehicle (MAV) and some kind of easily detectable markers on the floor. This setup was used for instance by the ETHZ Pixhawk team at the IMAV 2010 competition [4] or by all three winner teams of the autonomous Air Race at the RobotChallenge 2013. The drawback of this approach is that with limited ground clearance in indoor environments and limited camera aperture angle a lot of markers are necessary to cover larger flight areas. The precise deployment and the removal of the markers is time consuming and in the narrow time slots of a competition maybe not possible. The organizers of the RobotChallenge therefore provide a dashed figure eight line on the floor for their participants.

Another interesting solution was introduced by Kara in [5]. To determine the position, the MAV sends three slightly outward oriented laser beams down to the ground. The dots on the ground are captured by a fixed camera and a computer vision algorithm calculates the position which then is sent back to the MAV.

The Vicon MX motion capture system is a popular high end commercial solution offered by the UK based company Vicon. Universities like the Swiss ETHZ [6], [7] or the MIT Aerospace Controls Laboratory (RAVEN) use these systems for their development of autonomous aerial robots. The basic system consists of eight high speed, high resolution infrared cameras mounted on the ceiling of the flight area, a high speed network to connect the cameras and a central computer to merge the information from the cameras to attitude and position data for the observed MAV's. The camera lenses are surrounded by a ring of IR LED's illuminating the scene in front of the camera. On the MAV's four balls covered with a retro reflective coating reflect the light directly back to the cameras. The light passes through a daylight blocking filter and produces black pictures with bright white dots for the detected balls in the field of view. By combining the information from several cameras, position and attitude of several aircrafts can be computed with a rate of 200 Hz, a delay of 30 ms and centimetre precision.

A very interesting outdoor onboard computer vision system for relative localization in MAV swarms is presented in [8]. It is based on a Gumstix Overo computer board, a Gumstix Caspa camera module and an adapter board mounted on one MAV and weighing together 35 grams and a white board with a black ring as optical target on the other MAV. The ring detection is based on a blob detection algorithm to omit a time consuming Hough Transform. With full 752x480 resolution of the camera, a board size of 18x18 cm and an outer ring diameter of 14 cm a maximum range of 5.5 m is achieved. In tracking mode, where the approximate position of the ring and the threshold value for the greyscale to black/white conversion is known from the last frame, the frame rate is up to 27 fps. If tracking of position gets lost the whole frame has to be searched one time for the ring and the frame rate can drop to 7 fps. If also the threshold value gets lost the whole frame has to be searched several times with different threshold values until a valid ring is detected. This can be very time consuming. In the intended outdoor environments changes in light conditions occur slowly compared to the frame rate but in indoor environments the switching of additional room lights or at competitions a video camera spot or a photo camera flashlight can cause loss of threshold tracking and hence significant delays in localization.

2. System Concept

The computer vision system we had in mind on our survey should provide good value for money and it should be easily portable because at our university we can not afford a large lab with fixed installations dedicated to only one project. The Vicon solution needs a lot of permanently installed and calibrated cameras and costs several ten thousands of Dollars. Also the solution with the laser pointers on the MAV needs several cameras if used in a larger flight area and if the MAV rolls and yaws. The solution with markers on the floor and only one camera on the MAV is much more cost effective but has also the drawback of the necessary marker deployment and removal. When using a single marker on a wall or a post, similar to the above mentioned system detecting the black rings on white background, the setup of a lab can be done quickly but fast changes in light conditions keep a risk.

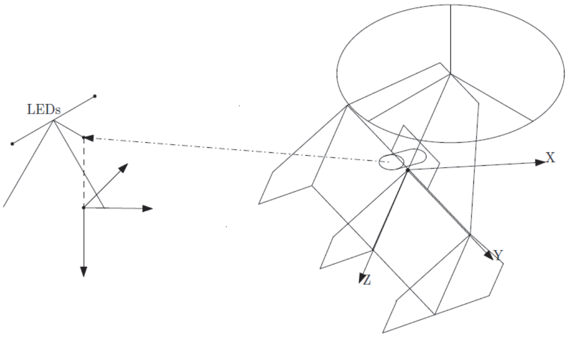


Fig. 1. Setup with a computer vision equipped MAV and a tripod with position lights in front of it

Our approach is now to use one small onboard CMOS camera that the hovering MAV directs towards three strong wide angle infrared LED's mounted on a tripod. By using LED's instead of a marker the system is independent of light conditions and the detection gets much easier. The tripod is placed best in a corner of the rectangular flight area so that nearly the whole flight area can be covered. Figure 1 shows the concept with the tripod in front of a single rotor hovering MAV.

The tripod (see Figure 3) holds three arms directed to the left, the right and to the front in a horizontal plane. Each arm is 50 cm long and holds a wide angle high power infrared LED at the end. If the infrared camera is exactly in front of the tripod centre it sees three equally spaced lights in line. If the camera is moved to the right, the middle light also moves towards the right light and vice versa. If the camera is moved above the LED plane the middle LED is below the connection line of the left and right LED and also vice versa. Because in Figure 3 the middle LED is near the right LED and below the connection line of the left and right LED it is clear that the camera is left above the tripod. The distance to the tripod can simply be deduced by the distance of the left and right LED on the camera picture. The more the camera is moved away the smaller the LED triple gets.

With this concept it is possible to measure the exact position of the camera relative to the LED's in a wide area in front of the tripod. The MAV has only to take care to point the camera roughly towards the LED's so that all three are in the field of view of the camera. For all kinds of hovering aircrafts this should be no problem. In our case attitude and heading is measured by a 9-DOF MEMS inertial navigation system. If necessary the flight area can be equipped with several sets of LED's so that the MAV is freer in its movements.

3. Key Components

For the infrared LED's OSRAM SFH4235 are used. They are high efficiency and high power IR LED's with a wide half angle of +/- 60 degrees. The maximum continuous current is 1 A at 3 V which produces an optical flow of typically 950 mW at 860 nm. Figure 2 shows the SFH4235 connected to the aluminum square tube arm for sufficient cooling.

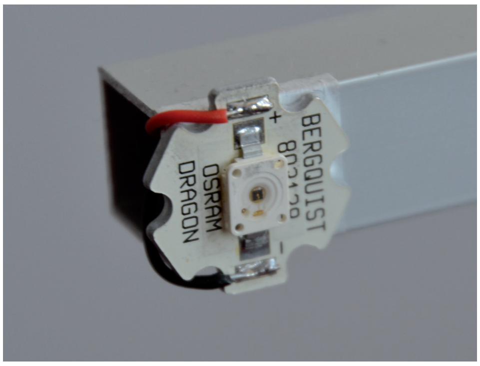


Fig. 2. OSRAM SFH4235 high power infrared emitter

Figure 3 shows the tripod on which the LED arms are mounted. The LED arms and the tripod legs can be folded so that the equipment can be stored space saving. If needed it is setup in minutes and can cover even very large labs.

Fig. 3. Tripod with IR-LED's

The CMOS camera used is a Gumstix Caspa FS for full spectrum imaging priced at 75 USD and weighing 23 grams. It features a 1/3-Inch Wide-VGA CMOS Digital Image Sensor and comes with a 3,6mm lens, aperture 2,0 and the angle of view is 94.91° / 68.46°. The sensor is an Aptina MT9V032 and it is capable of taking 60 pictures per second at a resolution of 752 x 480. Figure 4 shows the camera with an additional daylight blocking filter made of unexposed Fujichrome color reversal film that increases the contrast of the pictures significantly. The Fujichrome film is glued on a 3D printed lens adapter.

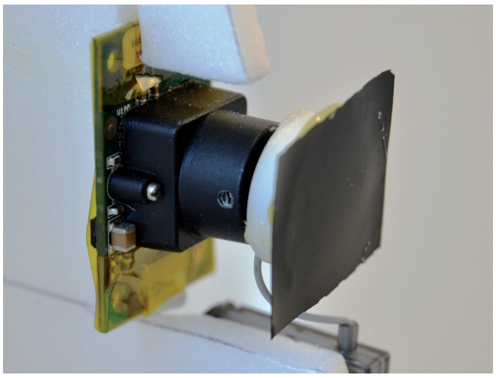


Fig. 4. CMOS camera with daylight blocking filter

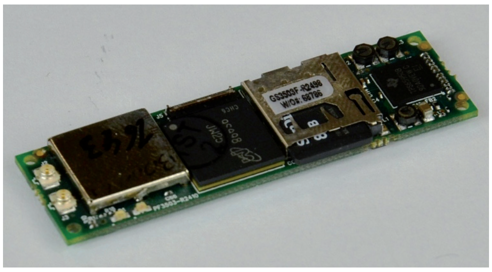


Fig. 5. Gumstix Overo Firestorm computer on module

The heart of the computer vision system is an Gumstix Overo Firestorm Com computer on module. It has a Texas Instruments DM3730 SOC that is typically used for mobile devices. It offers camera interfaces, a powerful 1 GHz ARM® Cortex™-A8 Core as a general purpose processor and a Video DSP. The whole Gumstix computer weighs only 6 grams and is priced at 189 USD.

Figure 6 shows the complete avionic system of our MAV. The CMOS camera with the Gumstix computer and a custom adapter board with a buck converter and connectors for the Gumstix is on the left side, the middle board holds the 9-DOF Inertial Measurement Unit and the board on the right is a ublox LEA-6T GPS receiver with a helix antenna. The bottom board below the three plug on boards holds the ARM Cortex-M3 avionik processor that controls the servos and motors of the aircraft, processes the IMU and position data and handles the communication.

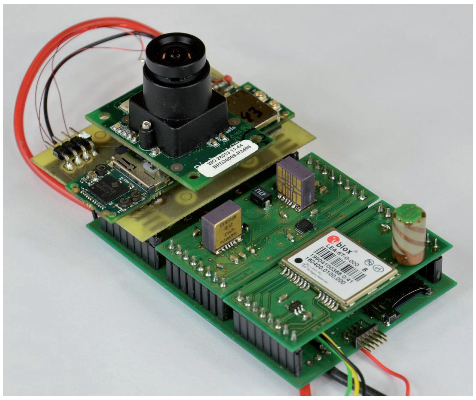


Fig. 6. Avionik system with Computer Vision, Inertial Navigation System and GPS receiver

For better handling and connectivity the Gumstix Tobi expansion board is used for software development and education (figure 7). However due to its weight and size it is not suitable for the MAV where we use our own circuit board. The expansion board has connectors for HDMI, 10/100baseT Ethernet, USB and a 40-pin header with GPIO, PWM and A/D lines. For protection and better handling all electronic parts and the camera are mounted in an ABS-casing.

Fig. 7. Development setup for the Gumstix computer with Tobi expansion board and Caspa camera

Because all parts used in our system are low cost of the shelf parts the total cost for the computer vision system is less than 350 $.

Table 1. Costs of embedded camera system

Item	Price[$]
Gumstix Overo	189
Tobi Base Board	69
Caspa Camera	75
Flex ribbon cable	5
ABS-Casing	5
Total	343

4. Operating System, Software and Tools

The operating system running on the Gumstix is an embedded Linux distribution called Ångström. It is loaded from the SD-card. The custom built image has already all required tools and libraries installed. The boot parameters passed from u-boot to the kernel are also stored on the SD-card in the boot.scr file. This means that a student receives a Gumstix computer, a Tobi baseboard and a memory card with the image and is up and running in five minutes. The operating system uses preconfigured static IP address and the user can connect using SSH over a direct Ethernet connection. This makes the system very flexible. In the process of developing programs for a UAV a computer with Ubuntu is used and the video output is streamed over the network to the connected PC. This setting offers a convenient work environment with a minimum effort, once the image is configured and compiled.

Due to the processing power of the processor, all applications are written and compiled directly on the target. For the development of the programs that are executed on the Gumstix Geany is used. It is a mall and lightweight integrated development environment (IDE) that is easy to use and runs on the target. This makes it the system of choice for compiling and executing the code.

For image processing an open source library called Open Source Computer Vision (OpenCV) is used. OpenCV is a free-to-use library for academic use. The library offers a wide range of algorithms for video and image processing. Today more than 2500 optimized algorithms are published with the OpenCV library. Our signal processing is based heavily on OpenCV. It also supports different programming languages and provides interfaces in C, C++, Java and Python. The library is widely used and perfect for use in education. Documentations and examples are easy to find [9].

The implementation of a highly reliable computer vision system is one of the main goals of this project. We achieve this by increasing the contrast of the LED's to their ambience to a very high level and by applying several filters in the image processing. To suppress unwanted daylight we use a daylight suppression filter made of an unexposed Fujichrome color reversal film to pass only infrared light to the camera lens. The infrared LED's, that are tracked, are also extremely bright and have therefore a high contrast to their environment. Figure 8 demonstrates an image with a long exposure time and the filter. The environment can still be seen in dark shades, the LED's are heavily overexposed.

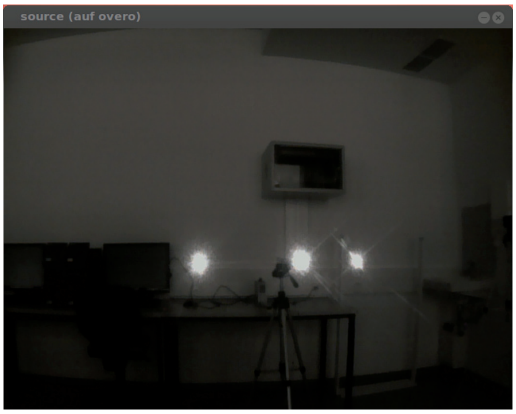


Fig. 8. Image with long exposure time and filter

Fig. 9. Image with short exposure time and filter

By further manual reduction of the exposure time we got the picture in Figure 9. With this setting also disturbances by light bulbs, fluorescent lamps or sunlight are suppressed very well.

As can be seen it is easy to track the objects now. The short exposure time also influences the size of the dots. They are smaller when the exposure time is shorter and the detected position in the picture is more precise. Another great advantage of very short exposure times is the robustness against blur because of camera movement. All our pictures are crispy sharp now, even if the MAV flies fast turns.

Many CMOS cameras include infrared blocking filters. For our project a camera without such a filter is necessary. We use the Gumstix Caspa camera. It is available in two versions, the Caspa VL (visible light) and the Caspa FS (full spectrum) which is used here in conjunction with the Fujifilm daylight suppression filter.

5. Image Processing

The OpenCV library provides a function for loading the picture from a memory of a lower layer. The first step of the signal processing chain is to transform the loaded color image to a grayscale image. Setting a threshold and converting the picture to black and white reduces the information further. A main characteristic of our method is that the contrast of the white LED spots to the nearly black background is very high. Selecting the threshold value is done by a histogram analysis. A function of OpenCV finds contours in the black and white picture. For human visual inspection circles are placed over these contours. The plausibility check is necessary to find the three correct objects that are searched. To find the correct position of these points a restitution of the coordinates is necessary. Finally the position determination can be calculated by using a 3-point resection. Before sending the information to the aerial vehicle the position is checked if the value is probable and possible. An overview of the image processing is given in Figure 10.

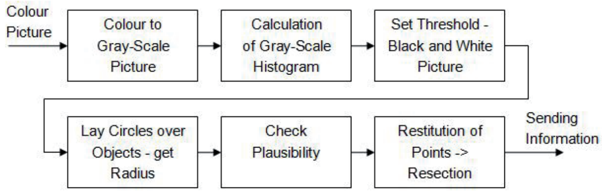


Figure 10. Overview of the image processing

Histograms are often used to analyse images. A histogram for a greyscale picture shows the brightness values on the x axis and the numbers of pixels that have this value on the y axis. The sum of the y values is the number of pixels in the image. This characteristic can be used for setting the threshold value. The infrared LED's are shown as white objects in the greyscale image. The rest of the image is dark or nearly black. Figure 11 shows how a typical histogram in our setup looks like. On the left side of the x axis is 0 (black) and on the right 255 (white). The upper histogram shows the complete distribution of the pixel values. On the left side the large number of dark pixels can be seen and on the right the view white ones. The lower histogram shows the distribution of the pixels zoomed into the white pixel area. In this area a peak can be detected, which is caused by the infrared LED's. They are assumed to be the brightest object in the picture. The threshold value is set shortly below this peak.

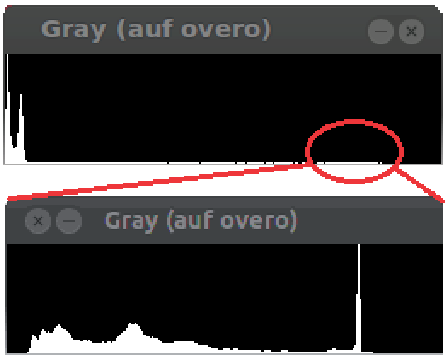


Fig. 11. Full histogram and histogram detail for bright pixels

If an error exists, because of sun, lights or reflections, it is necessary to check the plausibility of the detected contours. Filters and check algorithms raise the reliability to find the correct objects. An easy, but very useful filter is to check the radius of the objects. If a radius is too large or small the objects can be dropped. Another attribute of the objects is the similarity of their radii. The tracked objects get sorted by their radii. The three correct objects should be located near each other in this array. An algorithm calculates the average radius of the actual three objects. After that the maximal difference between the average radius and the radius of each object is calculated. This operation is done for every combination in the array. If five objects are tracked, the operation has to be done three times. When the operation is finished for every combination, the combination with the smallest difference is selected.

A significant contribution to the image processing time is caused by the correction of lens distortion. A standard algorithm corrects the distortion of the complete image before it is analyzed. In our system the white LED spots can be detected and measured in the distorted picture because they are very small. Only the resulting centre coordinates are undistorted. OpenCV provides the method cvUndistortPoints for distortion correction of individual pixels. This cuts down the time for lens distortion correction to a negligible value.

The problem of calculating the camera position and orientation out of at least three projected reference points on the picture, the camera parameters and the spatial coordinates of the reference points is called the three point pose estimation problem (P3P problem) or in photogrammetry the spatial resection problem. There are a number of direct and iterative ways for the solution with the first published already in the 19th century [10]. A well known algorithm that is also implemented in OpenCV is POSIT.

The iterative method used here was developed as part of a master's thesis [11] at our department. Compared to POSIT its main advantage is that it is much easier to understand for our electronics students with no special background in mathematics or computer vision. Their main focus in the control theory course is the control of multicopters and the vision system for the positioning should be as simple as possible. Another advantage is that our system uses only three LED's where POSIT needs four. The disadvantages are, that the roll angle is limited +/- 90° to get non-ambiguous results and that the calculation is not as effective as POSIT. But this does not matter as the lion's share of processing time is spent for picture analysis.

Figure 12 shows a simplified 2D setup. A, B and C are the three infrared LED's and N is the position of the lens center of the camera. The three angles between the three points seen from N are α (ANB), β (BNC) and γ (ANC). In a 2D scenario $\alpha + \beta$ is γ whereas in a 3D scenario $\alpha + \beta$ is larger than γ. The sequence of the three points on the picture plane from left to right has to be A, B and C, so that on the x axis of the picture B always lies between A and C.

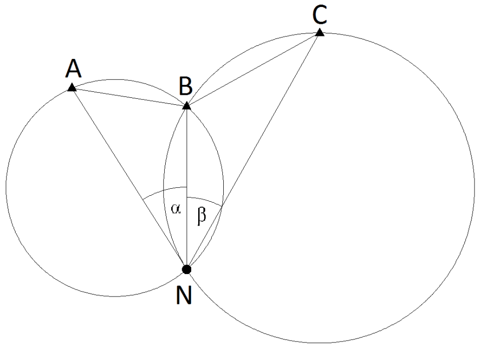


Fig. 12. Resection in 2D with the known points A, B and C

In a first step the three angles α, β and γ are calculated out of the coordinates A'B'C' of their projections on the picture plane. The picture coordinates are undistorted using the camera calibration data (OpenCV chessboard camera calibration) before that.

The side lengths of the projected triangle A'B'C' are calculated by formula 1. This is done in 2D with the positions of A'B'C' in the projection plane of the camera.

$$\bar{s} = \sqrt{(y_2 - y_1)^2 + (x_2 - x_1)^2} \qquad (1)$$

The distance from the camera center N to the projection plane is the focus length f. So the Euclidean distance from N to, for instance, A' can be derived by formula 2.

$$\overline{NA'} = \sqrt{A'_x * A'_x + A'_y * A'_y + f * f} \qquad (2)$$

With all lengths between N, A', B' and C' known, the angles α, β and γ can be derived with the law of cosines (formula 3) like in formula 4.

$$c^2 = a^2 + b^2 - 2ab * \cos\gamma \qquad (3)$$

$$\gamma = \cos^{-1}\left(\frac{a^2 + b^2 - c^2}{ab}\right) \qquad (4)$$

A remaining problem with α, β and γ is, that the angles are the same when B is above the plane ANC or below it. For that reason a set of alternative angles φ_1, φ_2 and φ_3 is used by defining a base point X' between A' and C' so that the connection from A' to C' is normal to the connection from X' to B'. The three alternative angles are then defined between A' and B' (φ_1), A' and X' (φ_2), and between X' and B' (φ_3). The sign of φ_3 is now holding the information if B is above or below the plane ANC.

The second step in our resection method is to iteratively assume a camera position n that gives the same angles φ_1, φ_2 and φ_3 as they were measured with the camera picture at camera position N. It is easy to derive the three angles because all positions (n, A, B, C) are known. The starting point n_0 is set some meters in front of the infrared LED's ABC.

The approximation technique that is used is the inexact Newton's method. Formula 5 gives an example for the one dimensional case. The iterations x_n are here scalar values.

$$x_{n+1} = x_n - \frac{1}{f'(x_n)} f(x_n) \qquad (5)$$

Because we have a three dimensional problem the Jacobian matrix is used instead of the first derivative to calculate the three dimensional iterations x_n in formula 6. At the beginning of the approximation it can happen that the inverse of the derivative gives too large values. Because of that, the factor p ($0 < p <= 1$) reduces this value in the first four iterations of the algorithm.

$$x_{n+1} = x_n - p * \left(J(x_n)\right)^{-1} f(x_n) \qquad (6)$$

After 10 iterations a position is estimated that is nearly correct. In Figure 13 a simulation of our spatial resection method is demonstrated. The red dots symbolize the infrared LED's, which are the reference points. The position of them is known. The green dots are the iterative steps to the correct position of the camera which is shown as a black dot. The start position for the iterative algorithm in this case is [0 0 0]. The last iteration steps cannot be seen because they are inside the black camera dot.

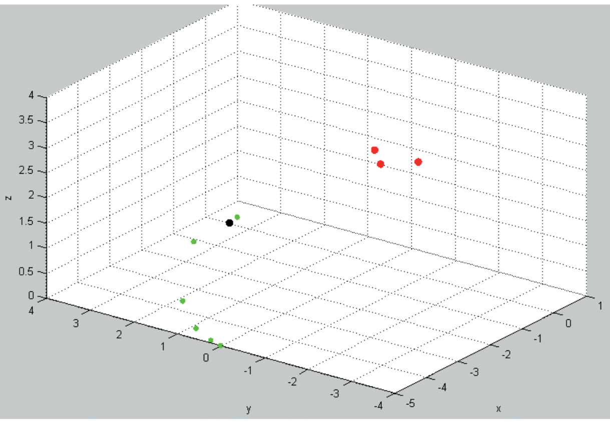
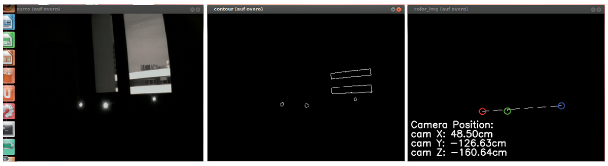


Fig. 13. Matlab simulation for iterative camera position estimation

6. Results

The computer vision system of the MAV project provides a low budget solution with good performance. With a frame rate of 12 pictures per second and a delay of only 130 ms the computer vision system covers a wide range of applications. In indoor environments the reliability was nearly 100 %. Of course a difficult environment can influence the reliability significantly. For indoor use it provides sufficient performance characteristics. In Figure 14 a difficult case for the computer vision system is shown. The left image shows the original image made by the camera with a short exposure time and the filter. This picture shows the three IR LED's in front of a window.

The adjacent building is in direct bright sunlight at noon. The image in the middle represents the contours of the objects that were identified. The right image shows the correct results. The plausibility check detects the correct objects and the position information is ready for sending to the navigation computer.



```
Camera Position:
cam X: 48.50cm
cam Y: −126.63cm
cam Z: −160.64cm
```

Fig. 14. Camera view and picture analyses

The achievable range of the system is limited with the exposure time. The distance between aerial vehicle and LED's was successfully tested at 10 m. This distance can be expanded to more than 10 m by choosing a longer exposure time but with reduced detection reliability. In this project a high reliability is more important than a long range.

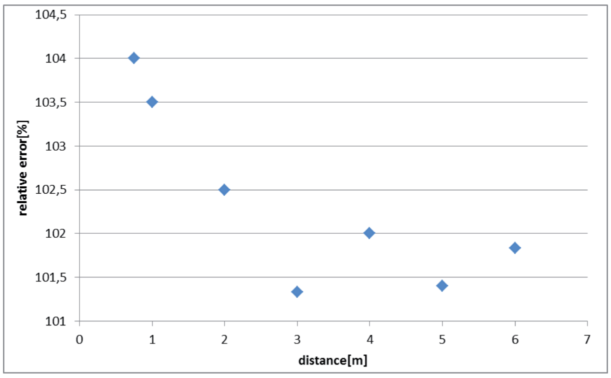


Fig. 15. Precision of computer vision range finder

The accuracy was tested at different distances. The results in Figure 15 show, that the imprecision is very small. Also at small distances the failure is not more than 2–3%.

7. Conclusion

Although the computer vision system was originally developed to work on board of a MAV it has also great potential for other projects with educational focus. Its reliability and precession combined with its small dimensions and low weight makes it the ideal choice for aerial vehicles. The used hardware and software makes it very easy to write programs once an OS image is created with the required tools and software. The system has also a high potential for educational use. The startup is well supported and also the OpenCV library is widespread. OpenCV knowledge can also be used for many other projects in the area of image processing.

ACKNOWLEDGEMENTS

This project is founded by the City of Vienna, department MA23, under grant "Endowed Professorship Embedded Control and Navigation, 10-06".

AUTHORS

Thomas Kittenberger*, **A. Ferner**, **R. Scheikl** – Department of Embedded Systems, University of Applied Sciences Technikum Wien, Vienna, Austria.

*Correspondig author. E-mail: thomas.kittenberger@technikum-wien.at

REFERENCES

[1] G. Freudenberger, *Sensor Fusion für ein MEMS-basiertes Trägheits-navigationssystem*, Master's thesis, University of Applied Sciences Technikum Wien, 2012.

[2] N. Vavra, Guidance, *Navigation and Control of a Micro Air Vehicle*, Master's thesis, University of Applied Sciences Technikum Wien, 2012.

[3] RobotChallenge Organization Team, "RobotChallenge 2013 Air Race Rules", http://www.robotchallenge.org/competition/

[4] L. Meier, P. Tanskanen, F. Fraundorfer, M. Pollefeys, "PIXHAWK: A System for Autonomous Flight using Onboard Computer Vision." In: *IEEE International Conference on Robotics and Automation*, 2011. DOI: http://dx.doi.org/10.1109/ICRA.2011.5980229

[5] M. Kara, S. Patra, A. Lanzon, "Designing Simple Indoor Navigation System for UAVs." In: *19th Mediterranean Conference on Control and Automation*, 2011.

[6] G. Ducard, "Autonomous quadrotor flight using a vision system and accommodating frames misalignment." In: *Industrial Embedded Systems, IEEE International Symposium*, 2009. DOI: http://dx.doi.org/10.1109/SIES.2009.5196224

[7] S. Lupashin, "The Flying Machine Arena as of 2010." In: *IEEE International Conference on Robotics and Automation*, 2011. DOI: http://dx.doi.org/10.1109/ICRA.2011.5980308

[8] J. Faigl, T. Krajnik, J. Chudoba, M. Saska, "Low-Cost Embedded System for Relative Localization in Robotic Swarms." In: *IEEE International Conference on Robotics and Automation, Karlsruhe*, 2013. DOI: http://dx.doi.org/10.1109/ICRA.2013.6630694

[9] I. Culjak, "A brief introduction to OpenCV." In: MIPRO. Proceedings of the 35th International Convention, 2012.

[10] R. Haralick, C. Lee, K. Ottenberg, M. Nölle, "Review and Analysis of Solutions of the Three Point Perspective Pose Estimation Problem", *International Journal of Computer Vision*, vol. 13, no. 3, 1994, Kluwer Academic Publishers, 331–356. DOI: http://dx.doi.org/10.1007/BF02028352

[11] R. Zarits, *Kamerabasierte Positionsbestimmung und Speicherung der Bilddaten auf einem Micro Air Vehicle*, Master's thesis, University of Applied Sciences Technikum Wien, 2012.