Preprints of the 2012 20th Mediterranean Conference on Control & Automation (MED), Barcelona, Spain, July 3-6, 2012

3D Position and Attitude Estimation Using Novel Marker Detection Method

Gergely Regula, István Gőzse and Alexandros Soumelidis

Abstract— **The aim of this paper is to present a novel indoor positioning system which is based on an image display method. The basic idea of the indoor positioning method is presented. The optical sensing is a combination of specific analogue preprocessing and digital signal processing techniques that are capable of measuring and recognising markers that are mounted on the agent to be tracked. The paper mainly focuses on the three-dimensional position and attitude reconstruction methods for multiple makers in known configuration. The estimation of a single marker's position is typically based on more than two sensors' measurements. Therefore, it is different from the classical epipolar geometry based methods. Information about the configuration of multiple markers is included in the reconstruction method as constraints.**

Index Terms— **Indoor positioning, multi-view 3D reconstruction, position & attitude estimation.**

I. Introduction

Recently, intensive research has been carried out in the field of individual and cooperative control of unmanned vehicles, and the experimental validation of the control methods and algorithms that are developed gain crucial significance. For this purpose, small-scale model vehicles are commonly used. An important part of a model platform suitable for this purpose is a positioning system that is capable of determining the position and attitude of the vehicles. Camera based motion capture (MOCAP) systems are most commonly applied to solve this problem. These systems – besides their advantages – suffer from several drawbacks, e.g. they usually require a dedicated laboratory with controlled ambient light. Furthermore, a large amount of computational power is required for processing the images provided by the cameras, as well as special optics, which are used for correct imaging. These all make camera based MOCAP systems rather expensive and inconvenient both in use and during development. A solution that has all the advantages of an optical MOCAP system, while being flexible enough to be realised at low cost, would be advantageous in the field of research of unmanned vehicles.

The approach proposed in this paper seems to fulfil the above requirements. A transparent graphic display acting as a variable pinhole camera is used as motion capture sensor to detect the position of one or more lighting markers. After a short presentation of the operating principle, the three-dimensional position and attitude reconstruction methods will be highlighted.

All authors are with the Systems and Control Lab, Computer and Automation Research Institute, Hungarian Academy of Sciences, 13-17 Kende u., Budapest 1111, Hungary. {regula,gozse,soumelidis}@sztaki.hu

II. Operating Principle

The image display method is based upon a setup consisting of three main parts: the lighting markers, the sensor and a TFT display. Markers are mounted on the model vehicle that is the subject of positioning. LEDs act as active markers. Their small size and low weight make them ideal for use on small-sized model vehicles. In most cases, the sensing element of an active marker is a camera. However, marker recognition involves complicated image processing. To avoid this problem, the proposed system uses phototransistor based optical sensing. By this means, a reliable link can be established between the LED marker and the sensing system. Reliability can be achieved by proper preamplification and signal processing. Since the marker's light is characterised by the frequency of the PWM signal it is modulated by, proper signal processing allows us to suppress the disturbing ambient light while preserving signals required for tracking the marker.

The key element of the whole system is the TFT unit, which makes the position estimation possible by the displayed picture in the image plane. Fundamentally, a TFT display consists of two main parts: the TFT layer and the backlight unit. If black and white are the only colours used in the image plane of the TFT display, the TFT unit can be interpreted as a special window that can be made transparent or opaque according to the image in it. This special window (TFT unit) is mounted with a distance of about 10 cm in front of a phototransistor. This way, if only a small-sized patch around the intersection of the TFT's plane and the line connecting the phototransistor and the LED marker is transparent, then the coordinates of the patch specify the direction of the marker relative to the phototransistor. If the patch can track the marker's motion, the instrument can provide direction measurements periodically. The spatial position of a marker can be reconstructed by geometric methods based on measurements of multiple sensors. The operating principle of the system is illustrated in Fig. 1.

A suitable indoor positioning system can track multiple markers. This system can accomplish multiple marker tracking by the use of several patches and independent LED recognition (each marker's light is modulated by different frequency).

The interested reader is referred to [1] for further details on the sensing method and the operation of a single sensor unit.

III. Hardware Setup

The structure of the whole positioning system is quite flexible. The positioning system consists of a number of individual sensor units and a PC. The sensor units are

978-1-4673-2529-5/12/$31.00 ©2012 IEEE

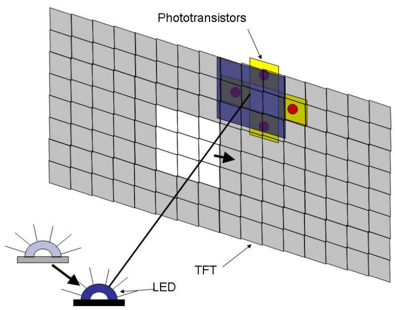

Fig. 1. Operating principle of the sensor units.

mounted onto fixed positions so that each segment of the space of interest is in the field of vision of at least two units. However, more than two increases the accuracy and performance of the system and it is also beneficial since markers might be covered by obstacles and other moving agents.

The sensor units are connected to the PC via CAN bus, which provides the required bandwidth. There are two modes of operation:

- The PC collects the coordinates of each marker's position on the image plane of every single TFT. In this setup, it is the sensor units' task to perform the tracking of each marker in their image planes.
- Using the collected preprocessed measurement data, it is the PC's task to calculate where the markers are most likely to be in the sensor units' image planes and to send these pieces of information to the sensor units.

Naturally, in both cases, the PC is responsible for performing the three-dimensional reconstruction at each time instant. The current operating frequency is set to 50 Hz.

The first mode handles hardware resources better (CAN bus load, processing power on the sensor units), while the latter can be utilised better during the development process.

Presently, the reconstruction algorithm is implemented on a Windows-based high priority application. In a future version, the PC might be replaced by a powerful embedded computer with a real-time operating system, to make the positioning system more compact and to make its operation more predictable.

IV. Estimating a Single Marker's Position

At present, our configuration consists of four sensor units mounted on the upper corners of the laboratory, pointing to the centre. To achieve better coverage, a fifth one is planned, which will be mounted on the centre of the ceiling, pointing to the ground.

The three-dimensional reconstruction method is fairly simple for a single marker. One of the greatest advantages of the sensing method can be utilised, namely that the so-called correspondence problem does not exist here. Therefore, given that a marker is visible to a number of the sensor units, its spatial position can be estimated by a least squares method.

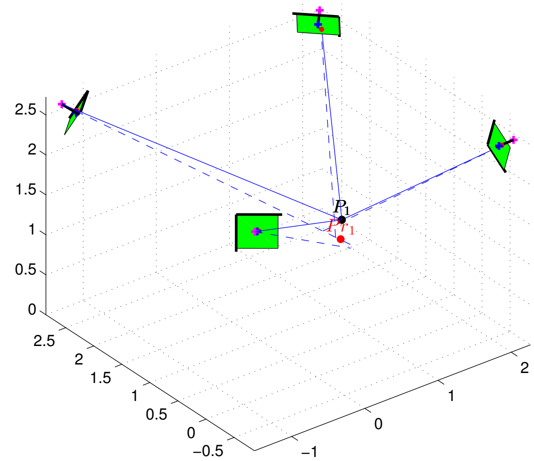

Fig. 2. Sensor setup and projection rays to the image planes.

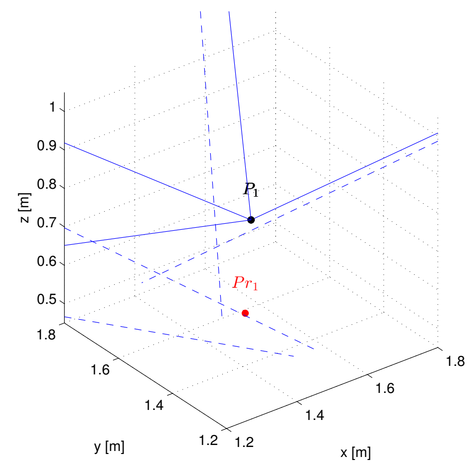

Fig. 3. Projection rays around the marker.

The optimal estimate of the marker's position minimises the sum of squared distances from the projecting rays. The method is illustrated in Fig. 2 and Fig. 3. Sensors are represented by their focal points (magenta), image planes (green rectangles) and axes. The size of the sensors is increased for better visibility. A straight line represents their viewing directions and the camera centres are marked by blue points. Solid lines connect a spatial point and the focal points of the sensors. The intersection of the projection rays and the image planes are quantized to pixels so that the real projection rays and those used for reconstruction (dashed lines) are not identical. In practice, due to the nature of the tracking method, these values are not necessarily integers as they are calculated from the sizes of the transparent patches.

The idea behind the reconstruction algorithm is illustrated in Fig. 4. The real spatial position of spatial point and focal point are marked by vectors p and f. The measured and real projections to the image plane are marked by black and red, respectively. From the origin, r points to the former.

Vectors v and w define a parallelogram, whose area can be

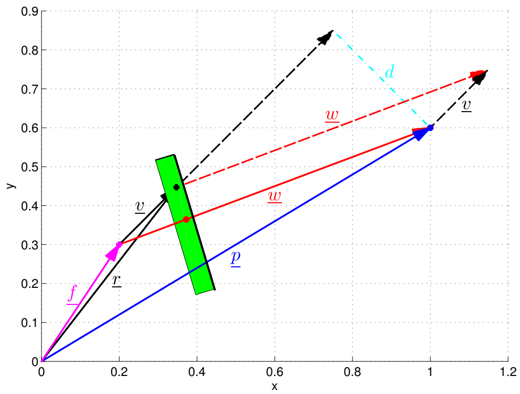

Fig. 4. Geometric explanation of position reconstruction.

calculated as follows:

$$|(\underbrace{\underline{r}-\underline{f}}_{\underline{v}}) \times (\underbrace{\underline{p}-\underline{f}}_{\underline{w}})| = d|\underbrace{\underline{r}-\underline{f}}_{\underline{v}}| \ , \tag{1}$$

where d is the distance between the real spatial point and the projection ray. The cross product can be written in matrix form $[\underline{v}\times] = \underline{\underline{V}}$. In the following, matrix and vector notations are omitted, unless they are not clear from the context. Matrices are denoted by upper case, vectors and scalars are denoted by lower case letters, respectively.
The squared distance can be written in the following form:

$$\frac{w^T V^T V w}{v^T v} = d^2 \ . \tag{2}$$

During the reconstruction step, the length of w is the unknown parameter. The cost function and optimal estimation are defined as

$$f(p) = \frac{1}{2}\sum_{i=1}^{N_s} d_i^2 \tag{3}$$

$$d_i^2 = \frac{(p-f_i)^T V_i^T V_i (p-f_i)}{v_i^T v_i} \tag{4}$$

$$p_{opt} = \arg\min_p f(p) \ , \tag{5}$$

where N_s is the number of the sensor units. The problem is a standard least squares problem, whose solution can be found by equating the cost function's derivative to zero.

$$\begin{aligned} \left.\frac{\partial f(p)}{\partial p}\right|_{p=p_{opt}} &= \sum_{i=1}^{N_s} \frac{V_i^T V_i (p_{opt}-f_i)}{v_i^T v_i} = \\ &= \sum_{i=1}^{N_s} \frac{V_i^T V_i p_{opt} - V_i^T w_i}{v_i^T v_i} = 0 \ . \end{aligned} \tag{6}$$

The optimal estimation can be extracted:

$$p_{opt} = \left[\sum_{i=1}^{N_s} \frac{V_i^T V_i}{v_i^T v_i}\right]^{-1} \left(\sum_{i=1}^{N_s} \frac{V_i^T w_i}{v_i^T v_i}\right) \ . \tag{7}$$

The matrix to be inverted has full rank if the configuration is not ill-posed, which is mostly the case due to the sensor setup.

A further refinement method is plausible. The reliability of a measurement can be quantified by the area of the transparent pixels on each sensor module. Therefore, it is good practice weighting measurements according to their relative reliability. The cost function has to be modified slightly:

$$f(p) = \frac{1}{2}\sum_{i=1}^{N_s} c_i d_i^2 \ , \tag{8}$$

where the sum of the weights is

$$\sum_{i=1}^{N_s} c_i = 1 \ . \tag{9}$$

This method slightly increases the accuracy of the reconstruction, which is currently about 1 cm when the marker is stationary.
The quality of measurements can be improved. Currently, due to the nature of the tracking of a marker, there is a theoretical limit on the allowable speed in directions parallel to the image plane of a sensor unit. However, smart setup can increase this speed without the need of modification of low level functionality. Suppose that an agent travels at a speed in a certain direction so that a sensor unit cannot track it. Then if the agent's position can be fairly well estimated without the measurements of this sensor, then projecting the estimated spatial point to its image plane can increase the speed of the convergence of the marker tracking. This can be performed side by side with the reconstruction algorithm without significantly increasing the computational demand on the PC.
It is worth mentioning that the described method differs from the most commonly applied reconstruction methods, which make use of epipolar geometry [2], [3]. Surprisingly, indoor positioning systems developed by research groups rarely utilise than two optical sensing units.
For further information, the reader is referred to [1], in which real-time operation of the system is presented in more detail.

V. Multiple Marker Based Attitude and Position Estimation

Depending on the requirements of the application, estimating an object's position only might not be sufficient. Therefore, several markers need to be mounted on the vehicle. Since the estimation of all the three attitude angles is the more general case and this is what is required in aerial vehicles' case, this will be discussed in the remainder of the section.
In general, the problem is as follows. Let K_V be the frame fixed to the centre of gravity of the vehicle and K_E the Earth fixed frame of reference. Suppose that the markers' coordinates are constant in K_V. The aim of attitude and position reconstruction is to find the transformation parameters between K_E and K_V.
Other notations in this section will also appear. The origins of K_E and K_V in their frame are denoted by O_E and O_V, respectively. Spatial points are denoted by capital letters, the vectors pointing to points are denoted by lower case letters.

It is straightforward that at least three markers are required to calculate the attitude of a flying vehicle. However, more markers are recommended, even though the computation requirement increases, since their use can increase the robustness of the method.
The basic idea of attitude and position estimation is that a cost function is defined that penalises the squared distances between all the markers and the corresponding projection rays (as previously introduced), and additionally, the deviation of the estimated structure of the markers from the "known" is also included. The term "known" is important in the sense that the structure might only be known to a certain precision. If the mounting points of the markers are designed by the use of a CAD program, then the structure is well defined. However, there is always the possibility to mount them randomly (in a sensible way). The distinction is important since they yield somewhat different methods that will be made clear. It is worth mentioning that even imprecise information on structure might significantly simplify attitude estimation since the reconstructed structure will be the same as the expected one.

A. Exact Information Based Structure Reconstruction

If mounting points of all the markers are known exactly (e.g. by CAD design), then it is reasonable adding structure information to the optimisation problem as equation constraints. Constraints should be as simple as possible in order not to make the problem extremely complex. Such constraints could be distances between pairs of markers or angles between two edges connecting markers, etc. Since purely angle information specify similar objects of different sizes, distance information is more useful and has the same complexity. Therefore, mostly these will be considered in practice.
The problem is formulated as follows. Let the cost function f be the aggregate penalty for each marker as in (5) [1]

$$f = \sum_{m=1}^{N_m} f_m = \frac{1}{2} \sum_{m=1}^{N_m} \sum_{i=1}^{N_s} d_{m,i}^2 \ , \qquad (10)$$

where N_m is the number of markers and $d_{m,i}$ is the distance of marker m and the corresponding projection ray of sensor i.
As hinted above, the domain of the optimisation is restricted to a subspace of $\mathbf{R}^{3N_m}$ defined by a set of equation type constraints. The problem can be solved by the Lagrange multiplier method, which is the well-known special case of the Karush-Kuhn-Tucker approach in nonlinear optimisation. Several advanced algorithms exist (e.g. interior point methods) that can solve these types of problems, many of which are available as callable software libraries.

B. Structure Reconstruction Based on Approximate Information

One way of the problem formulation is as follows. Let f_M be the aggregate penalty for each marker defined the similarly as in (10).
Structure information can be included in various ways. The simplest method is to measure the distances between a number of marker pairs and include and additional term in the cost function that penalises the deviation from these values:

$$f_S = \sum_{(i,j) \in N_s} \left[r^2(p_i, p_j) - r_0^2(p_i, p_j) \right]^2 \ . \qquad (11)$$

Here N_s denotes the set that includes the marker pairs (i, j) the distance between which is measured. Distance can also be measured more reliably than angles, therefore this kind of additional penalty seems to be the most reasonable among all the possibilities.
The total cost function is the weighted sum of the two above,

$$f = f_M + c f_S \ , \qquad (12)$$

where the coefficient c is a design parameter.
The problem above is a fourth order unconstrained optimisation problem, which is more complex than the one introduced in Section IV. Since no direct solution exists in this case, either, iterative solution involving search is still needed. Not only the order of the problem makes it complex, but also its dimension, which is $3N_m$. As a consequence, the required operating frequency limits the maximum number of markers.
The reason why the structure of the markers is not included in the optimisation problem is the following. Since these pieces of information contain uncertainty, it might happen that the constraints could not be satisfied at the same time. As a consequence, the result might be far from ideal.
It can be seen that the inclusion of more distances increases the computational costs of the calculation of both the cost functions and the derivatives. Therefore, for computation power reasons, it is good practice including only the fewest possible distance measurements that may not lead to a different structure, keeping in mind that shadowing might occur.

C. Extracting Position and Attitude Information

Ideally, a homogeneous transformation defines the relative position and attitude of the world frame and the vehicle fixed frame. This description is commonly used in robotics and computer graphics. The transformation matrix is composed of a rotation and a translation part as

$$T = \begin{bmatrix} R & p \\ 0^T & 1 \end{bmatrix} = \begin{bmatrix} I & p \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} R & 0 \\ 0^T & 1 \end{bmatrix} \ . \qquad (13)$$

However, it is not good practice calculating this transformation matrix directly (e.g. by a least squares method) for several reasons. One of them is the possible loss of precision in case of large distance from the world frame's origin, and another is that the rotation part has special structure that the result should possess. Hence, decomposing the problem is beneficial. As a first step, the position information should be extracted, followed by the attitude angles.
It is worth mentioning that in most cases, none of the markers can be placed exactly in O_V, which adds extra calculation

[1] The sum of the weighted cost functions in (8)'s yield slightly better results.

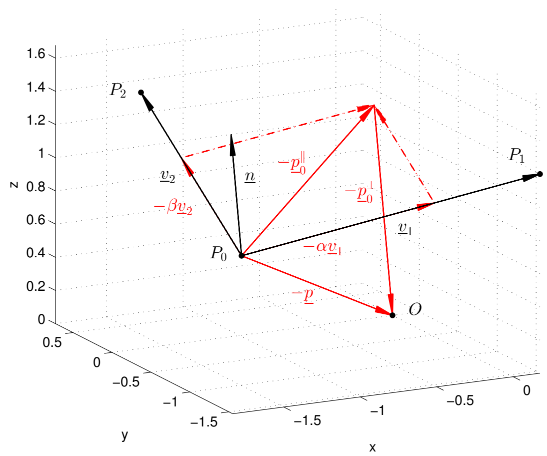

Fig. 5. Determining the position of O_V

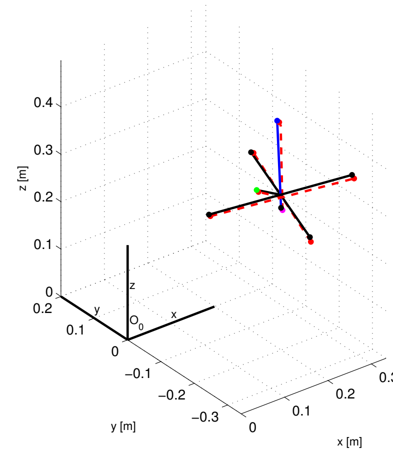

Fig. 6. Marker setup and reconstruction.

steps to the reconstruction algorithm.

If the spatial alignment of the markers is known exactly, three of the markers uniquely defines the position of O_V in K_E. By simple vector algebraic manipulations, the origin's coordinates can be found. A possible method is as follows. Let p_0, p_1 and p_2 be the vectors pointing to three markers in K_V that uniquely span the plane $\mathcal{P}$. Let $v_1 = p_1 - p_0$ and $v_2 = p_2 - p_0$. The vector p_0 is a linear combination of v_1, v_2 and n, which is the normal vector of $\mathcal{P}$.

$$p_0 = p_0^{\|} + p_0^{\perp} \tag{14}$$

with the notations

$$\begin{aligned} p_0^{\|} &= \alpha v_1 + \beta v_2 \\ p_0^{\perp} &= z n \ , \end{aligned} \tag{15}$$

where

$$\begin{aligned} z &= p_0^T n \\ n &= \frac{v_1 \times v_2}{\|v_1 \times v_2\|} \end{aligned} \tag{16}$$

Since the values of z, α and β are known in advance, the relative position of O_V from p_0 can easily be calculated. The method is illustrated in Fig. 5.

Slightly more complex is the case when the reconstructed spatial alignment of the markers does not match the model exactly. An approach similar to the one in Section V-A can be applied, i.e. penalising the deviation of the estimated position of O_V from the expected. The computational requirements for this step is not significant compared to that of the structure reconstruction step.

Attitude estimation is quite straightforward then. First, the O_V (and all the markers) should be shifted so that $O_E = O_V$. Then, the rotation part R of T can be determined by a least squares method.

Two issues might arise, tough. Firstly, rank deficiency might cause problems, which can be tackled by adding a proper virtual marker to the points. Secondly, the resulting matrix might not hold the properties of a rotation matrix, i.e. its eigenvalues are not of unit length. However, since in practice the length does not differ significantly from 1, it can be corrected by e.g. eigenvalue decomposition.

Depending on which attitude representation is required, the angles can be extracted as desired [4], [5].

D. Practical Considerations and Implementation

A crucial step in nonlinear optimisation is finding an appropriate starting point. It can dramatically reduce the time required for reaching the optimum. Fortunately, a reasonable starting point can easily be found, simply by disregarding the structure constraints. As seen in Section IV, this reduces the problem to an LS one, which can be calculated independently for each marker and requires little time.

The first experimental application of our attitude reconstruction method will aid the control of a small-sized indoor unmanned quadrotor helicopter. The helicopter will be equipped with four markers placed in the edges of a square. The diagonal of the square is 20 cm and its centre lies 3 cm above the centre of gravity. Mounting four makers on the airframe is sensible both for symmetry and robustness reasons. As an illustration, the helicopter and the results of an example reconstruction is also included in Fig. 6, where the black parts correspond to the "real" vehicle and the estimation is coloured red. The green point shows the direction of the helicopter and the virtual marker is coloured blue.

As the airframe is completely designed in a CAD program, it is reasonable applying the reconstruction method presented in Section V-A, paired with the simpler vector algebraic position calculation and attitude extraction methods in Section V-C in the final solution.

Before (quasi-)real-time implementation, the methods were tested in synthetic environment. MATLAB with Optimization Toolbox were required for initial simulations. The setup of the sensor units can be modelled in a virtual environment. All the required parameters of the sensor units are readily available, since a precise calibration procedure has already been carried out (based on [6] and [7]). Real measurement

Env.	Method	Δt_{mar}	Δt_{tot}	e_{pos}	e_{att}
MATLAB	Lagr	84.52 ms	85.88 ms	0.28 cm	0.8 deg
MATLAB	WD	81.96 ms	95.88 ms	0.28 cm	0.8 deg
C	Lagr	-	10.00 ms	-	-

TABLE I

PERFORMANCE COMPARISON BETWEEN THE PROPOSED METHODS.

data are not yet available, as the tracking of multiple markers is not solved yet. Thus, measurements need to be generated by projecting the markers onto the image planes of the TFTs. Measurement errors and calibration imperfections can also be modelled.

MATLAB also provides great flexibility during the evaluation and fine tuning of the algorithms. It is also important examining which method suits the best together. It turned out during simulation steps that all the structure reconstruction methods could be improved by adding extra constraints/penalties. Even though including all the distance information between the markers uniquely define the alignment, the four markers never lie in the same plane after the optimisation, not even in case of constraint optimisation. An additional constraint, i.e. the midpoints of the diagonals of the square should coincide, improves the quality of reconstruction. Its effect turned out to be significant in case of relatively large weight. In case of constrained optimisation the reason is the steeper derivative in certain search directions.

The idea that including an additional virtual marker improves the calculation of the matrix R describing the attitude of the vehicle was motivated by the fact that this symmetric structure always results in rank deficient matrices.

Despite its advantages, MATLAB has a major drawback, i.e. performance. An efficient program performing the same algorithm may perform 10 – 100 times better, depending on the platform and the nature of the algorithm, especially if it makes use of GPU acceleration. Based on MATLAB simulations though, the relative efficiency of the proposed algorithms can be estimated before selecting the one to be implemented efficiently. Tab. I shows a comparison of average time required for different methods. The results are based on 2000 different position and attitude combinations, run on a 2.2 GHz Windows based PC. The reason for the choice of the operating system is the lack of driver available for the special CAN communication capable card that is used. The abbreviations “Lagr” and “WD” correspond to the constrained and unconstrained methods. It can be seen that the majority of time is spent on estimating the position of the markers (see Δt_{mar} and Δt_{tot}). As a comparison, the average time required for the whole process is shown in the last row, when the algorithm is coded efficiently. An increase of a magnitude can be seen in the performance, which confirms the expectations.

The method involving constrained optimisation is implemented as a program coded in C. The key to efficient implementation is a suitable nonlinear optimisation library. Such libraries are available. For simplicity, the student version of KNITRO was chosen as it is supported by the Optimization Toolbox. It is free of charge for a limited period. Therefore, in the future it is likely to be replaced by Ipopt, which is free of charge tool with similar interface, provides great flexibility, although some of its supported solvers are not free for commercial purposes.

In general, these optimisation libraries provide an interface through which the functions to be minimised, the constraint functions the derivatives and the Hessian can be evaluated. These functions have to be implemented by the user, which is rather tedious compared to an implementation in MATLAB. For efficient linear algebraic manipulations, BLAS and LAPACK routines are used and the libraries are compiled from source in order to achieve maximum performance. As a result, it is not the position and attitude estimation algorithm that limits the operating frequency of the positioning system.

VI. CONCLUSION AND FUTURE WORK

As mentioned earlier, a reliable multi-marker tracking method has not been implemented yet. This is the key element to attitude estimation, which makes it crucial. An efficient tracking method for a single marker is also important, it can still be improved, although it operates sufficiently well. Multi-marker tracking brings additional difficulties that have to be tackled. From the algorithmic point of view, the most relevant issue is the handling of overlapping transparent rectangles.

Several additional aspects can also be investigated. It is planned that the system will be made capable of tracking multiple vehicles. The risk of shadowing effect is more likely to have greater importance. Therefore, a suitable method for estimating the positions of shadowed markers might improve the robustness of the system.

The portability of the code is also of great importance as the optimal solution is implementing the algorithm on a powerful real-time platform, thus eliminating the need of a PC.

ACKNOWLEDGEMENT

This work is partly supported by the Control Engineering Research Group of HAS at Budapest University of Technology and Economics.

REFERENCES

[1] G. Regula, I. Gőzse, and A. Soumelidis, “Position estimation using novel calibrated indoor positioning system,” in *Proc. IEEE International Instrumentation and Measurement Technology Conference*, 2012, To Appear.

[2] O. Faugeras, “From geometry to variational calculus: Theory and applications of three-dimensional vision,” in *Proc. IEEE and ATR Workshop Computer Vision for Virtual Reality Based Human Communications*, 1998, pp. 52–71.

[3] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.

[4] M. D. Shuster, “A survey of attitude representations,” *Journal of the Astronautical Sciences*, vol. 41, no. 4, pp. 439–517, 1993.

[5] J. Diebel, “Representing attitude: Euler angles, unit quaternions, and rotation vectors,” Stanford University, Tech. Rep., 2006.

[6] R. Duraiswami. (2000) Camera calibration. Electronic. University of Maryland - Institute for Advanced Computer Studies.

[7] R. Tsai, “A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses,” *IEEE Journal of Robotics and Automation*, vol. 3, no. 4, pp. 323–344, 1987.