Three-dimensional object recognition using a monoscopic camera system and CAD geometry information

Alessandro Cefalu*, Jan Boehm
Institute for Photogrammetry, Universitaet Stuttgart,
Geschwister-Scholl-Straße 24D, 70174 Stuttgart, Germany

ABSTRACT

Current research conducted by the Institute for Photogrammetry at the Universitaet Stuttgart aims at the determination of a cylinder head's pose by using a single monochromatic camera. The work is related to the industrial project RoboMAP, where the recognition's result will be used as initiating information for a robot to position other sensors over the cylinder head. For this purpose a commercially available view-based algorithm is applied, which itself needs the object's geometry as a-priori information. We describe the general functionality of the approach and present the results of our latest experiments. The results we achieved show that the accuracy as well as the processing time suite the project's requirements very well, if the image acquisition is prepared properly.

Keywords: object recognition, machine vision, optical inspection, quality assurance

1. INTRODUCTION

The increasing competition in industrial production forces manufacturers to tailor their products according to individual customer needs. Hence, the manufacturers either produce a broad range of products or a large variety of a few base products. In order to produce small lot sizes at reasonable costs, the production facilities have to be used at their optimum. This leads to a flexible production scheme, where different products are processed on the same production line. The same flexibility is required from any inspection systems, which is to be placed within such a production line. In-line inspection is an essential part of production quality assurance. Ideally, 100% of all products and features are tested. The earlier a defect can be detected, the fewer are the costs, since fewer resources are wasted in needless subsequent production steps. Strict quality assurance is an important distinctive feature of premium sector manufacturers and secures long-term investments.

Considering the afore-mentioned factors, we can specify the demands on an industrial inspection system as follows:

- It needs to be fast, to test as many features as possible without slowing down the production process.

- It needs to be flexible, to measure different variations of a product.

- It has to operate directly in the production line, to detect defects as early as possible.

- It has to be precise and robust, to guarantee high quality standards.

The use of flexible optical sensors provides the ideal solution to these demands in a versatile production environment. Already numerous optical gauging and inspection systems have established themselves in industrial practice. The broad variety of optical measurement methods has created an even broader range of systems and suppliers. We wish to mention only two complementary methods from this broad range, namely triangulation and interferometry. Structured white light triangulation is a proven technique in industrial inspection and is continuously used in many different applications. Such systems can be successfully used directly in the production line due to their high speed of data acquisition and still maintain their high precision [1]. Structured white light triangulation is especially suited to measure the outer, i.e. convex, geometry of an object. An attribute that is typical for triangulating systems. In contrast using interferometry, we can implement measurement systems, which are ideally suited to measure concave geometry, such as the inside of a drill hole [2].

*alessandro.cefalu@ifp.uni-stuttgart.de; phone +49 711 685 83397; fax +49 711 685 83327; www.ifp.uni-stuttgart.de

Fig. 1 Picture taken during a test at the University of Applied Sciences of Augsburg, which is one of our project partners. This successful joint experiment was performed in order to test data exchange between our camera system and the robot control unit.

It is current state-of-the-art to operate these sensors in a stationary manner. I.e. the sensors are in a fixed position towards a certain part or a feature thereof. However, if a different variant of the part has to be tested, the sensor might have to be repositioned. This creates the need to reconfigure the inspection system, which in turn creates additional costs or causes downtime. Therefore, it is highly desirable to automate the reconfiguration procedure. In early approaches specially designed measurement machines were employed, which integrated several sensors to inspect a part [3]. Such specialized machinery not only is expensive but also lacks scalability and flexibility. Later developments use industrial robots for senor handling. For applications with low to medium accuracy requirements such systems have already entered the industrial market [4]. For high-accuracy applications such a combination of multi-sensor measurement and industrial robotics is a current scientific and industrial challenge.

The work described within this paper is part of a larger project dealing with the in-line inspection of a cylinder head. The partners of this project entitled RoboMAP combine their expertise in industrial robotics and optical measurement to build an in-line inspection system, which satisfies the needs of a modern production environment as described above. In this environment, the accuracies required range from below 1µm to about 100µm. Therefore, it is unrealistic to assume one senor alone or a combination of sensors at the same accuracy level can provide the solution. A combination of several sensors at different accuracy levels is obviously required. It is common knowledge, that the higher the accuracy or resolution of a sensor is the smaller is its working range or field of view. Consequently, the positioning requirement of each involved sensor varies. The sensor with the highest accuracy has the tightest requirements on the positioning. Roughly, the requirements for positioning are two orders of magnitude above the sensors accuracy. To achieve suitable positioning at all accuracy levels, a coarse to fine strategy is employed. In this framework, each sensor has to maintain an accuracy that is suitable to derive positioning information for the next higher level of accuracy. At the very beginning of this chain an initial pose estimate has to be derived, which is accurate enough to steer the robot's movement and initiate the measurement. We achieve this initial pose estimate by using a camera placed above the work area and employing object recognition strategies. The specifications for this system require a 100% recognition rate over a working area of 0.8x0.8m in less than a second and an accuracy of the pose estimate of 1mm in translation.

2. VIEW-BASED OBJECT RECOGNITION

Literature describes two basic approaches of object recognition, a view-based approach and a feature-based approach [5]. In the latter image features, for example edges, are successively assigned to equivalent model features until a valid position in space can be retrieved from the combination of assignments. Hence, if the model consists of m features and the image contains n features, the quantity of possible solutions in an initiating step amounts to mn. Obviously the computational complexity of this approach is strongly dependent of the number of features and thus is not suited for the use with complex objects such as the cylinder head used in this work. Especially when the range of the object's position can be restricted, the view-based approach is more appealing.

To solve our task a view-based object recognition approach was chosen, which is provided by the image processing software HALCON [6]. It can coarsely be divided into two parts. First a set of artificial views is created. The set is referred to as 3D shape model and has to be computed only once in an off-line phase. This 3D shape model is then used during the actual object recognition.

In order to create the 3D shape model the object's geometry must be available in form of a CAD model. It is used to generate virtual two-dimensional views of the object, which are compared to the real image in the object recognition step. First a spherical shell around the object's center has to be defined. This area represents the possible range for a camera's position relative to the object and is specified in spherical coordinates. Additionally a range for the rotation of the camera around its optical axis, which always points to the object's center, has to be defined. Hence, eight parameters need to be declared, a range for longitude, latitude, radius and roll angle each. In this spherical shell an automatically defined number of camera stations is distributed. The CAD model is then projected into each virtual image plane using the real camera's calibration parameters to ensure correct mapping. In order to reduce processing time the virtual camera stations are sorted hierarchically and image pyramids, i.e. different resolutions, are used. The lowest hierarchy level contains the highest number of views with full resolution. Both, number of views and resolution decrease with increasing level.

Additionally, the view's creation can be influenced through various other parameters, e.g. maximum number of hierarchy levels or minimum face angle. The latter is used to remove edges from the virtual views which are contained by the CAD model but might not be visible in the real image. Especially model edges which approximate free-form surfaces fall into this category. If the angle between two adjacent faces is smaller than the defined threshold the corresponding edge is ignored in the view creation.

The actual recognition algorithm can be understood as an enhanced 2D matching. The quantity of virtual views is searched for the best 2D matches, beginning on the highest hierarchy level. Here a number of possible candidate branches of the hierarchy tree is identified. Each branch is then followed downwards by successively repeating the search on the child nodes of the previous best match until the lowest level is reached. Optionally the result, whose accuracy at this point is restricted by the discrete distribution of camera station, can be refined by estimation.

The algorithm may be applied to detect multiple objects of one kind. If the number of possible matches is limited by the user, the best matches are returned. The decision if a match is valid or not can be controlled through a parameter which can be interpreted as the fraction of edges in the view which match with the object's edges in the real image. Therefore, if the parameter is set correspondingly it is possible to detect objects even if they are partially occluded. On the other hand it can lead to a less robust recognition, especially if the object's edges are displayed with inadequate contrast. We also observed an increasing number of obviously wrong pose detections, when trying to recognize an object with simple geometry in images containing a lot of disturbing texture, as the correlation of edges becomes more ambiguous. At the same time a significant deceleration of the recognition process can be observed. Hence, such a parameter setting should be avoided, if possible.

3. EXPERIMENTAL STUDY

3.1 Data preparation, choice of equipment and workspace design

Many aspects influence the quality of the result and therefore should be considered during the planning phase. In most cases these influences are correlated so that one could think of them as a network, whose center is taken by the object of interest itself. The cylinder head is a demanding object for recognition purposes, not only because of its complex geometry but also because of its surface characteristics. It mainly consists of machined surfaces of an aluminum cast which are typically non planar free-form surfaces, boreholes, etc. Due to the rough machining of the surfaces in most parts, the reflection can be described as diffuse. However, in some parts there is fine surface finishing and we observe specular reflection. While the rough areas should be strongly illuminated in order to create high edge contrast, fine regions easily cause overexposure. In particular non planar areas with specular reflection are difficult to handle.

Furthermore the geometric complexity of the cylinder head leads to a very extensive CAD model. The processing time of the 3D shape model creation and of the object recognition would be far beyond our requirements, if the complete model was used. The model contains a lot of information on the inner geometry, which is of no use for the object recognition anyway. These circumstances led us to restrict the model to a single surface, the combustion chamber sealing face (CCSF). It is the largest reflecting surface and has distinct edges which define unambiguous features. As the face is planar it is possible to handle its reflectivity. In fact this circumstance is exploited in our workspace design as will be described later on.

For image acquisition we use a monochromatic uEye UI-2250-M camera, which has a resolution of 1600x1200 pixels and a pixelpitch of 4.4µm and a Schneider-Kreuznach lens with a focal length of 12mm. As mentioned before calibration of the camera system is crucial for successful object recognition as the creation of the 3D shape model depends on the calibration parameters. In order to obtain parameters corresponding to HALCON's calibration model, we performed the calibration using the functionalities provided by the software. These include the creation of a calibration pattern with circular marks and automated measurement of the marks in an image sequence. The estimation does not perform a free network adjustment and thus deformations caused during the printing of the pattern and its mounting on a solid platter are not accounted for. In order to deal with this deficiency we first performed a separate calibration with AUSTRALIS using the same image sequence. As AUSTRALIS computes a free network adjustment we obtained corrected coordinates of the marks. The original mark coordinates have then been replaced before performing the HALCON calibration whose results can be seen in Table 1. Equation (1) depicts the definition of the radial distortion parameter K as described in the HALCON reference documentation. Here u and v are the coordinates of an image point in the camera coordinate system and u' and v' are the corresponding coordinates after correction of the radial distortion.

Table 1. Calibration parameters as described in the HALCON calibration file and the corresponding residual

Parameter	Value	Unit	Description
Focus:	12.7479	[mm]	Focal length of the lens
K:	-970.805	[1/m²]	Radial distortion coefficient
Sx:	4.39842	[µm]	Width of a cell on the CCD-chip
Sy:	4.40000	[µm]	Height of a cell on the CCD-chip
Cx:	796.712	[pixel]	X-coordinate of the image center
Cy:	607.811	[pixel]	Y-coordinate of the image center
W:	1600	[pixel]	Width of the video images
H:	1200	[pixel]	Height of the video images
Residual:	0.0812612	[pixel]	

$$u' = \frac{2u}{1+\sqrt{1-4\mathrm{K}\left(u^2 + v^2\right)}}$$

$$v' = \frac{2v}{1+\sqrt{1-4\mathrm{K}\left(u^2 + v^2\right)}}$$

(1)

To determine the best position for the camera several calculations have been carried out, assuming possible positions of the cylinder head to be in a field of 0.8x0.8cm. Fitting this area into the image and assuring a low ground sampling distance, while leaving enough space for a robot to work at the same time resulted in positioning the camera at a distance of ca. 1.6m to the center of the area at an elevation of 60°. The optical axis was aligned in an angle of 30° with respect to the zenith.

Previous experiments showed that illumination has a strong impact on robustness and accuracy of the object recognition. Therefore camera settings such as exposure time and aperture have to be adapted correspondingly. After the camera setup was determined, our main goal concerning the lighting of the cylinder head was to use the CCSF's reflectivity in such a way that it would be displayed very brightly in an image while all other areas should remain dark, thus creating ideal edges in the images. This demanded a custom lighting solution as the cylinder head is no standard object concerning illumination. Most commercially available lighting solutions aim at the handling of small sized objects from a short distance and thus do not meet our requirements as our object is large in scale and again space for the robot to work must be maintained. In addition our solution has to be switchable with low latency as it will have to be shut off during the measurement of the robot sensors in the future. These constraints led us to the use of high intensity LED bars which can be triggered with very high frequency and intensity if necessary. We decided to choose red LEDs and a corresponding optical lens filter in order to make the images' quality more robust against fluctuations of other light sources. At this point we would like to mention that the use of the optical lens filter has a great influence on the camera parameters, especially on the focal length of the lens and the radial distortion coefficient. Therefore it is essential to perform the calibration with mounted filter.

To create homogenous lighting across the whole CCSF independent of its position in the workspace, without causing overexposure, a very diffuse lighting situation had to be established. We accomplished this by positioning the bars on the side of the workspace opposing the camera's position and pointing them to a white matt panel located above the workspace. By adjusting direction and intensity of the lighting as well as exposure time and aperture of the camera we managed to create a lighting solution which completely fulfills our intensions.
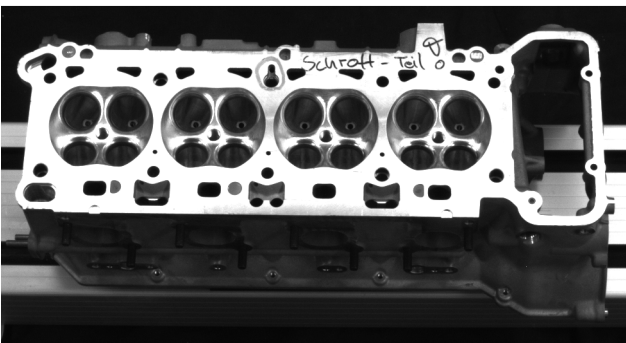
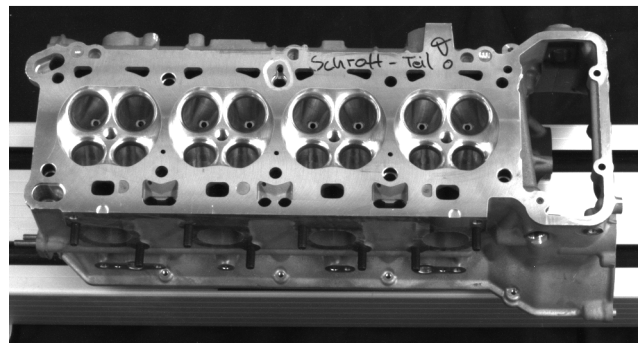Fig. 2. Image taken using indirect LED lighting

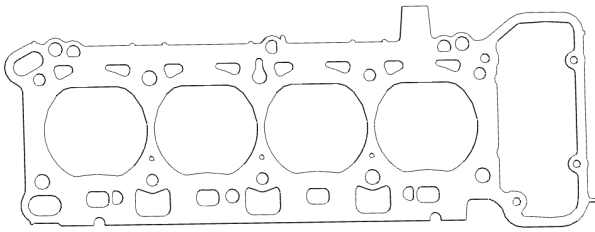Fig. 3. Image taken using ordinary ceiling lamps
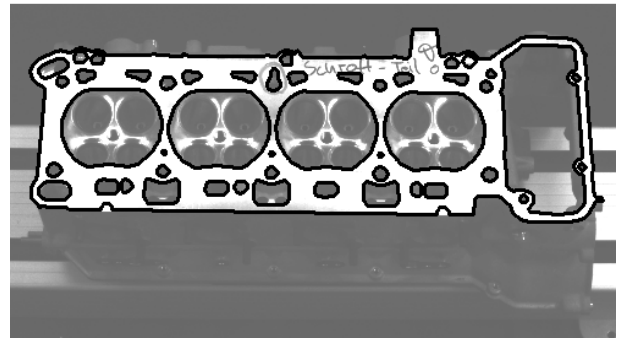
Fig. 4. CAD model after reducing it to the sole CCSF

Fig. 5. Desaturated image of the cylinder head with projection of the CCSF contours after successful object recognition

3.2 Testing and Evaluation

In order to test the approach's suitability for the RoboMAP scenario we simulated a small section of an assumed production line. We placed a linear unit, whose accuracy is specified with 10µm, inside the workspace and placed the cylinder head on its carriage. The carriage was then positioned on six stations with intervals of 5cm. On each station ten images were taken thus obtaining sixty images to perform the object recognition on. This way the pose results can be used to gain different accuracy information. While the ten results of each station yield information on the approach's repeatability, the known distances between the stations enable us to rate its absolute accuracy. In order to point out the importance of proper illumination the test was carried out twice; once using lighting as described above (figure 3) and once using the laboratory's ceiling light, ordinary fluorescent lamps (figure 4). In both cases a recognition rate of 100% was reached. This is a very important requirement to industrial applications. The maximum processing time for the images taken without LED lighting was 0.94s. This does not include the time needed for image acquisition. Additionally, we are planning to use a SmartCam, i.e. a camera with integrated CPU, in the final implementation of the system. This eliminates the need for a separate computer for the object recognition, but on the other hand a slower performance must be expected as its CPU capacity is not comparable. In contrast, the maximum time needed for images taken with LED lighting was 0.51s, which much more suits our requirements. The reason for this speed-up is the different parameter set for the object recognition, when compared to the case of unfavorable lighting conditions. The tweaking of parameters in the object recognition algorithm heavily affects run-time performance as mentioned above.

Precision: On each station the mean value and the standard deviation was computed for every degree of freedom, as shown in Fig. 6. One can see that the maximum deviation in the position expectedly is reached along the Z-axis, which corresponds to the camera's optical axis. However, this maximum of 0.078mm as well as the maximum deviation 0.025° in the orientation can be considered very satisfying. In comparison the test without LED lighting yielded maximum deviations of 9.296mm and 2.255° respectively.

Absolute accuracy: In order to retrieve information on the absolute accuracy we computed all combinations of relative transformations between the sixty detected poses, leaving out double combinations of course. This leads to 1770 possible translations with given ground truth for the translation vector. Initially rotational errors have no impact on the position of the origin of the coordinate system, which is at the object's center. However, it is obvious that they cause increasing errors on points with growing distance to the origin. In order to account for rotational errors, we determined the deviations for a point at the outer rim of the cylinder head, thus providing an upper bound to the error.
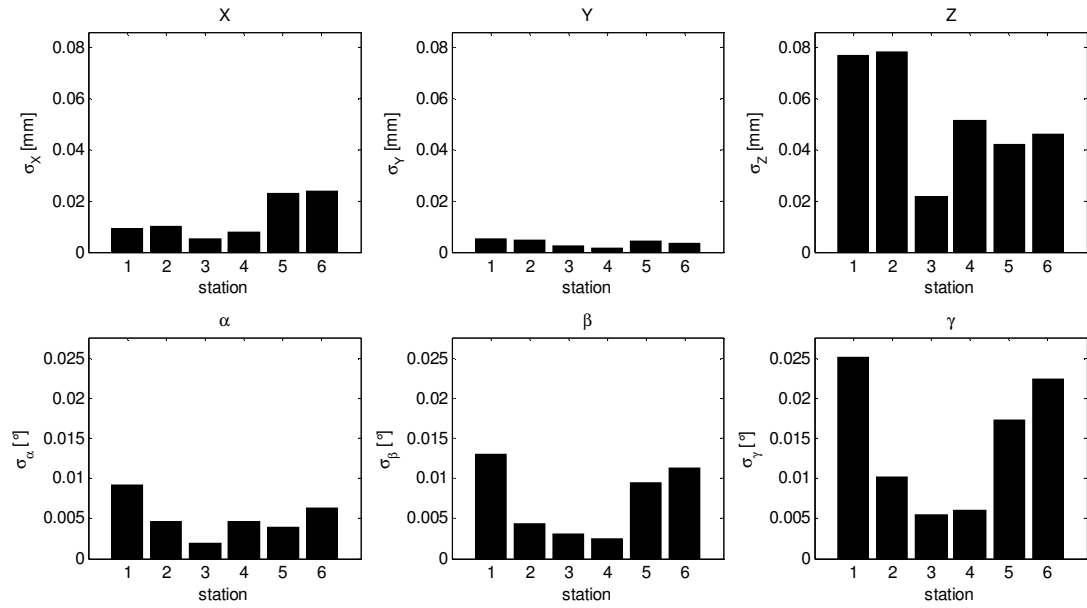
Fig. 6. Standard deviations of the six degrees of freedom, computed for every group of ten images at the six carriage positions.

Sorting the results eases interpretation and reveals information on systematic behavior, as it becomes more obvious when plotted on a graph. For example sorting by station number reveals a comparatively high error rate in relative poses involving the first station, as depicted in Fig. 7. This corresponds to the precision results at this station. If the results are sorted by the distance in-between stations, a clear growth of the errors' absolute value becomes observable, as Fig. 8 shows. Nevertheless the results clearly fulfill our project's demands. We obtained a total error span of 1.020mm with an over-all standard deviation of 0.204mm and a mean error of -0.150mm. In contrast the results for the tests without LED lighting show al total error span of 34.718mm.
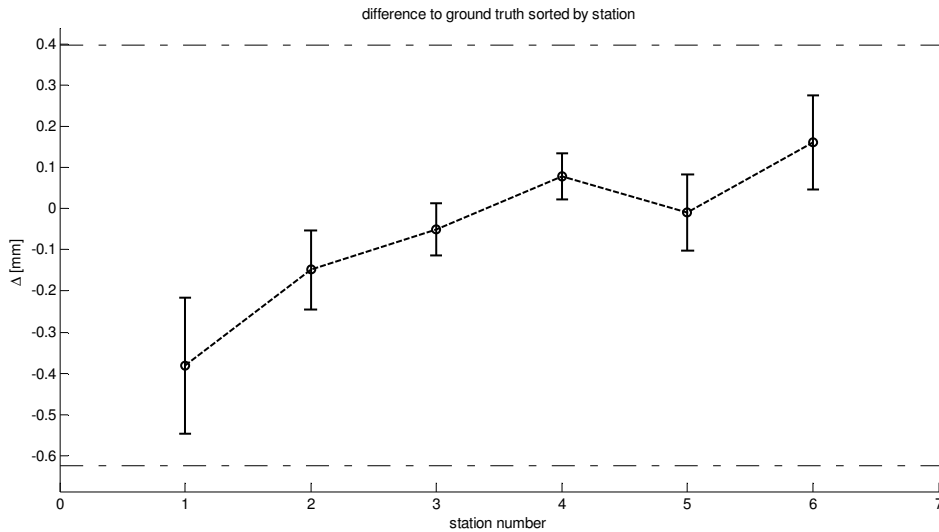

Fig. 7. Differences to the ground truth of the translation distance. Results were sorted by involved station. The circles represent the mean value. The vertical bars show the corresponding standard deviation. The upper and lower horizontal line represent the over-all maximum and minimum error.
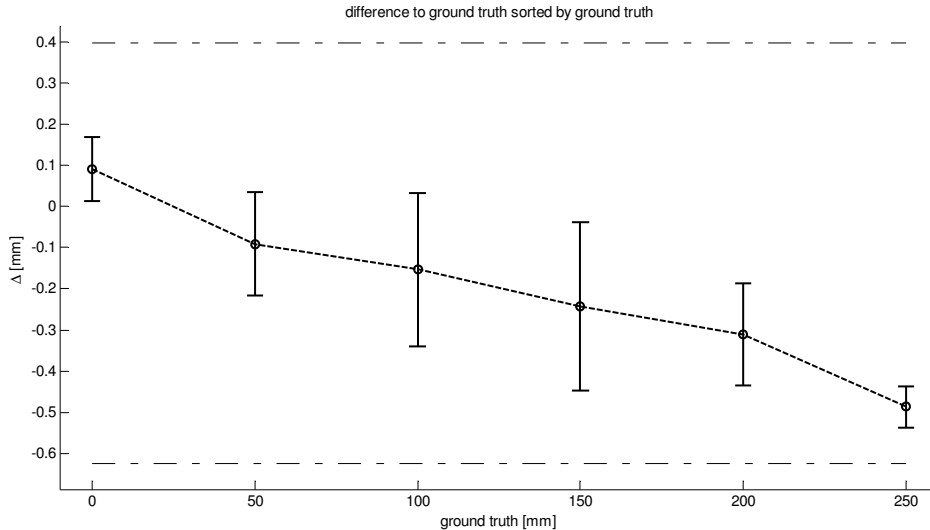
Fig. 8. Differences to the ground truth of the translation distance. Results were sorted by ground truth.

4. SUMMARY

The work described above is a successful application of state-of-the-art object recognition technology to a manufacturing and inspection scenario. We have shown that off-the-shelf sensors can be utilized at maximum measurement accuracy when proper calibration is carried out. By investing into quality illumination robustness can be ensured to a high degree even when the object is difficult in nature. We have shown this in practical experiments in a simulated production line.

While the results we have presented fulfill our initial requirements, we still see room for improvement. For example, the deviation of the pose clearly exhibits a drift. Additionally, we observe that all measurements involving the first pose show a significantly larger error. In future work we will investigate these cases and if we are able to eliminate them, the accuracy will increase even further.

Our research work will now focus on the coupling of the object recognition results to the robot's control unit and subsequent positioning of additional sensors. The combination of all sensor measurements and the robot control information in a common coordinate system is an additional challenge. Proper calibration of all sensors in themselves and relative to each other is indispensable and provides a further interesting field of application for photogrammetric technology.

5. REFERENCES

[1]   Frankowski, G., Chen, M., Knuth, T., "Optical measurement of the 3D coordinates and the combustion chamber volume of engine cylinder heads", Proc. Fringe 2001 (2001)
[2]   Knüttel, A, Rammrath, F., "Spectral coherence interferometry (SCI) for fast and rugged industrial applications", Lasers and Electro-Optics and the International Quantum Electronics Conference, CLEOE-IQEC 2007 (2007).
[3]   Boehm, J., Gühring, J., Brenner, C., "Implementation and calibration of a multi sensor measuring system", Proc. SPIE 4309 (2001).
[4]   Roos, E., "Method and device for calibrating robot measuring stations, manipulators and associated optical measurement devices", United States Patent 6615112
[5]   Grimson, W. E. L., "Object recognition by computer: The role of geometric constraint", The MIT Press, Cambridge (1990).
[6]   Wiedemann, C., Ulrich, M., Steger, C., "Recognition and tracking of 3D objects", Patern Recognition, Lecture Notes in Computer Science, Volume 5096 (2008)