3D Vision Sensing for Grasp Planning:
A New, Robust and Affordable
Structured Light Approach

Tilo Gockel, Johannes Ahlmann and Rüdiger Dillmann

IRF - Institute for Computer Design and Fault Tolerance
Universität Karlsruhe (TH)
D-76131 Karlsruhe, Germany
gockel@ira.uka.de

Pedram Azad

HRCN, CNS
ATR Computational Neuroscience Laboratories
Kyoto, 619-0288, Japan
pazad@atr.jp

Abstract— **In this paper we present a new approach to 3D shape acquisition. This implementation enables a robot arm to move the scan unit over the object without the scan unit being tracked. The chosen structured light approach uses an initially unknown white noise pattern that can easily be projected with any fixed pattern projection system. Object points are acquired on the basis of finding correspondences in the pattern in the camera image of the current scene. This is achieved by using a fast SSD correlation algorithm. In the current test setup we are using a standard video beamer and a standard digital camera, but we have also set up a scan head for fixation on a robots' wrist. The special requirements of this miniaturized system will be explained as well as our implementation. Our approach reduces hardware complexity to a minimum using only one calibrated camera and one calibrated fixed pattern projector.**

In this paper we present the whole system also including the calibration procedure. The focus will be on the hardware setup, but we also give an introduction to the software methods used.

Index Terms— **3D Model Acquisition, Projector Calibration, Pattern Projector, Structured Light.**

I. INTRODUCTION

3D-SCANNING in robotics normally requires large and heavy sensor mechanics that are not usable for real-time scanning of movable objects. In this paper we want to present a new approach using a one-shot acquisition system, that can easily be miniaturized. Furthermore a 3D registration on-the-fly can be realized using this system. The system is implemented using a fixed pattern projection-based scan mechanism, a fast online registration and a fast visualization of the resulting 3D data. Perhaps the most famous results of such an approach were published by Rusinkiewicz et al. [5]. These researchers used a multi-pattern projection system with a complex image-to-image stripe tracking mechanism. Due to the need of projecting various patterns over time this system lacks

lacks the possibility of using any fixed pattern projection system and therefore cannot be miniaturized. Approaches of other researchers using one-shot methods based on a specific pattern depend on a very precisely manufactured pattern slide [8-12]. To avoid these drawbacks we have developed a one-shot / single pattern projection system with no demands for accuracy of the slide pattern. Our system is based on the principle of finding correspondences of the pattern in the camera image of the current scene. Each correspondence provides one object point by triangulating between the projector and the camera. The test setup uses a standard video beamer projecting only one white noise pattern. Later on this projector will be substituted by a simpler device using a low-cost gobo known from multimedia applications as projection slide. The scene is observed by a standard digital camera.

Figure 1 shows the workflow we intend to realize using this new scan head. After 3D-scanning the object and meshing the resulting point cloud, a 3D registration with a collection of shape images from a library will be done.
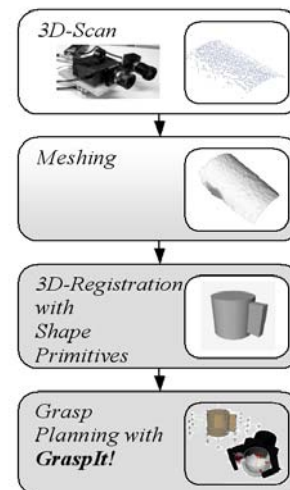

Fig. 1: Grasp Planning Workflow using *GraspIt!* [17,20]

Later on a specific grasp of the robot's hand can be selected using the software *GraspIt!* (focus of this paper is the implementation of the sensor in hard- and software. For details regarding *GraspIt!* please refer to [17,20]).

Normally projector calibration is done using a well-known pattern [1,8]. Our aim was to calibrate the pattern projector with the same pattern which is used during the scan process, without the need to be known to the system initially.

Section II of this paper explains the calibration procedure and the correlation and ICP algorithms we have developed. Section III shows the test-bed hardware and the software environment. In Section IV the results of the system are summarized, ending with a conclusion and future work in section V.

II. ALGORITHMS AND WORKFLOW

A. Camera and Projector Calibration

The scan system is calibrated in two steps. In the first step the camera is calibrated, which is then used for calibrating the projector in the second step.

For the calibration of the camera we use an extended DLT (*Direct Linear Transformation*) approach, which also deals with radial lens distortions [1,4]. This algorithm uses an iterative procedure in order to calculate the optimal set of parameters. In each iteration the DLT parameters are calculated first, incorporating the distortion parameters of the last iteration for de-skewing the camera image.

After this the parameters modeling the radial lens distortion are calculated using a least squares approach. [1] The input of this test-field-based calibration method consists of a set of point pairs (P_{World}, P_{Image}) where P_{World} is an a priori known 3D-point and P_{Image} the corresponding 2D-point captured by the CCD-chip. We acquire this set using a stack of acrylic plates of known thickness with the top plate containing a well-known dot pattern (cp. figure 5 in chapter III).
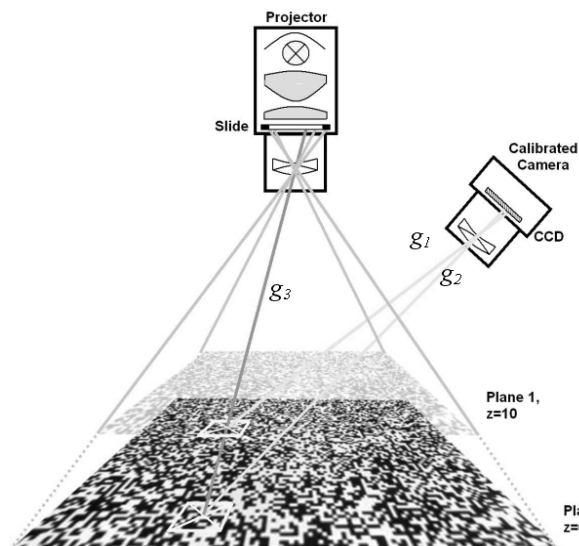

Fig. 2: Schematic sketch of the systems' geometrical layout used for Projector Calibration

Discrete 2D-points are acquired by determining the centers of the dots. Projector calibration is done utilizing the same extended DLT algorithm which is used for camera calibration (figure 2).

However the input for the algorithm is acquired in an entirely different way. Initially the pattern is projected onto the plane defined by $z=0$ and captured by the calibrated camera. This image will serve as a template for the whole following process. In the second step the pattern is projected onto planes in different heights ($z=h$) using the acrylic plates.

Now the same correlation algorithm which will be utilized during the scan process later on (except neglecting the epipolar constraint and searching in two dimensions) is used for finding correspondences of the template in the other takings. For each 2D-correspondence that is found the related 3D-points are calculated by using the parameters of the calibrated camera and intersecting the resulting camera rays with the well-known plane of height h. Naming the 3D-point in the template $P_{Template}(n_i,m_i,0)$ and the corresponding 3D-point $P_{z=h}(x_i,y_i,z_i)$ the resulting point pair for the input of the calibration algorithm is defined as ($P(x_i,y_i,z_i)$, $P'(n_i,m_i)$). Figure 2 shows the intersections of the camera rays with the projector rays illustrating the cooperation of the camera and the pattern projector.

We would like to mention that, with this method, it is not the projection of the original pattern is being calculated but the projection of the template image. Thereby, an object point is acquired during the scan process as follows: For each correspondence found, the camera ray g_1 related to the 2D-point in the camera image is calculated by using the parameters of the calibrated camera, and the camera ray g_2 related to the 2D-point in the template image is calculated in the same way.

Next, the 3D-point $P_{Template}(n,m,0)$ corresponding to the template match can be calculated by intersecting g_2 with the plane defined by $z=0$. Using the parameters of the calibrated projector one can now determine the projector ray g_3 easily by inserting the coordinates n and m.

Finally the object point is defined by the intersection of the straight lines g_1 and g_3. If no intersection can be found the closest point between g_1 and g_3 is used for the further calculations. [4,15,16]

B. Correlation Techniques

To retrieve local patterns in the camera image one has to calculate the correlation beween camera image and template image. We first tried an implementation of the Least Squares Correlation algorithm (LSC, [2,4,15]) which works well and which is even capable of dealing with perspective distortions but takes too much calculation time and therefore cannot be used in this real-time system.

As a result the current system uses a standard Sum of Squared Differences correlation technique (SSD, [13,16]) which can be implemented considerably faster and which delivers almost equally good results.

This SSD works as follows: Given a template image window with the gray value function $f(x,y)$, a search image window with the function $g(x,y)$, a window size of $2W+1$ and a pixel

$(i,j)^{\mathrm{T}}$ in the template image, the SSD correlation algorithm searches for the displacement vector $(d_1,d_2)^{\mathrm{T}}$ that minimizes the following error measure $c(d_1,d_2)$ [13,15,16]:

$$c(d_1,d_2) = \sum_{k=-W}^{W} \sum_{l=-W}^{W} (f(i+k,j+l) - g(i+k+d_1,j+1+d_2))^2$$

As a conclusion we want to mention that the strategy is to permit even bad matches as the result of the search for a pre-condition, since a bad match can be justified by perspective distortions at this stage. A further quantification is done using plausibility considerations like: "This matching pattern is in the near neighbourhood of the last one delivering a high correlation. Even if the current correlation is weak, it is nevertheless likely that it is a correct match". [15]

C. Fast 3D Registration

One important requirement for scanning of dynamic scenes on-the-fly (in real-time) is the ability to do a fast online registration from scan to scan. This is not the focus of this paper so we will only give a short insight: the current implementation uses an ICP (*Iterative Closest Point*) variant which utilizes the *vtkLandmarkTransform* class of the VTK (*Visualization Toolkit*) software library for the minimization of the error measure [6,7,14]. Special attention has to be paid to the closest-point-search, which is the bottleneck of the ICP algorithm. In the current implementation we are using a 2D grid structure to store the 3D points (figure 3). [15,16]
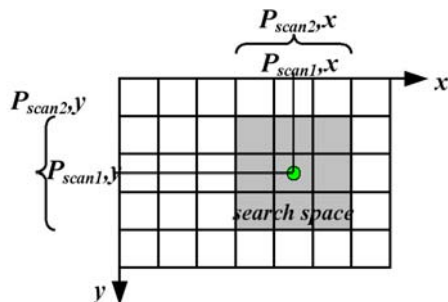

Fig. 3: Grid Structure for the Closest Point Search

P_{scan2} is indexing in the search space defined by P_{scan1}. This approach of accelerating the closest point search through storage in a 2D grid is allowed given the 2 1/2D-data type of our data but one has to mention that this implementation is only working with small angles of approx. 10-15° between scan *n* and scan *(n+1)*.

III. IMPLEMENTATION

A. Test-bed Hardware

In our first test-bed hardware we are using a standard video beamer for projection purposes. This enables us to switch easily different resolutions and different patterns. Later on in this section we will describe another projection method using a gobo projector with strobe lighting.

The camera in the test setup is a standard digital video camera which is connected to a standard BT878-based frame grabber. The system is assembled using a profile construction to allow easy adjustability in all directions. In the left corner of the base plate one can see a bracket, which serves as a stopper for the acrylic calibration plates (figure 4, 5).
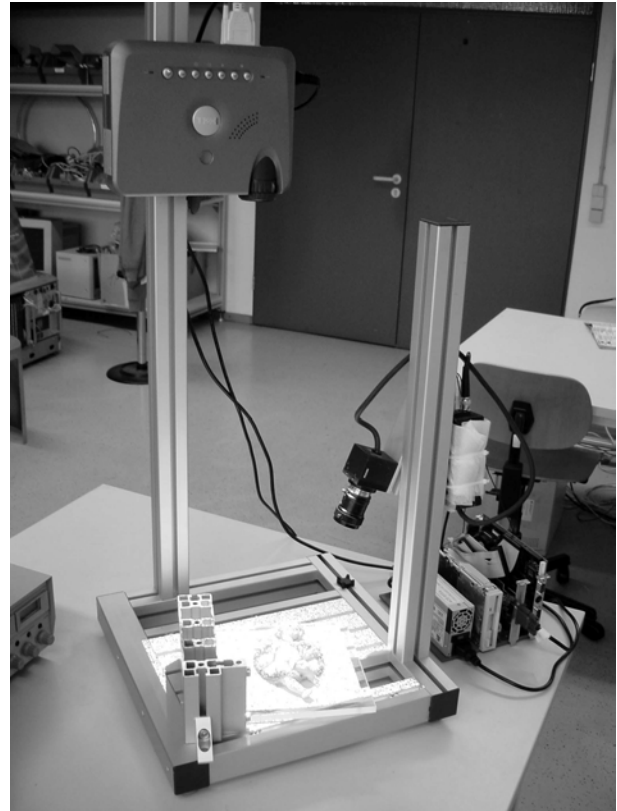

Fig. 4. Photo of the testbed hardware consisting of a standard miniature video beamer on top and camera on the right.

Since we are still experimenting with different projection patterns at the moment, a small embedded PC is necessary for this purpose, located on the right hand side of the test-bed.
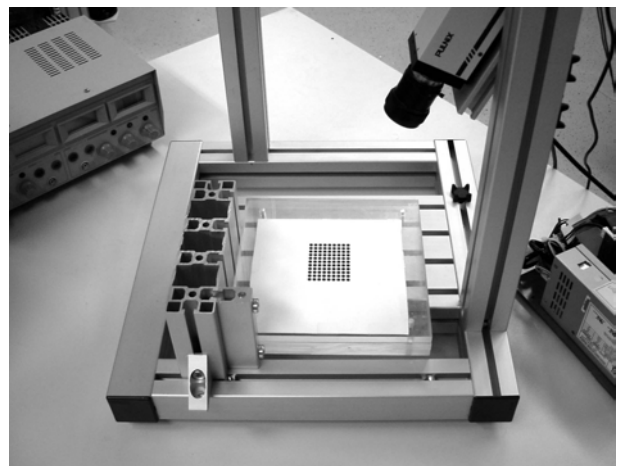

Fig. 5. Photo of the testbed hardware used for the camera calibration. A dot pattern is placed on the bracket stopper.

We want to mention that naturally it does not make any dif-

2531

ference whether the object is moved or the complete scan unit is moved or even both the object and the scan unit are moved at the same time.

B. Miniaturized Scan Unit on the Robot

For assembling this type of sensor on the wrist of our humanoid robot ALBERT [18] it had to be miniaturized. The pattern projector on this new scan unit was realized using a standard photo flash with a C-Mount lens mounted in front and a gobo (small chrome-on-glass slide known from multimedia applications) fixated in between. (figure 6,8)
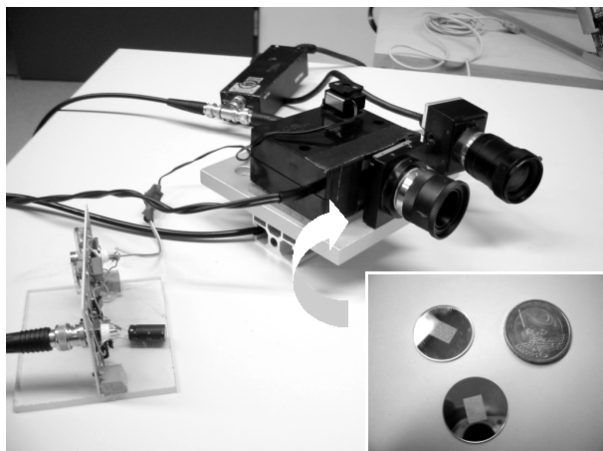

Fig. 6: Miniaturized Scan Unit and Gobo Slides and Sync Electronics

Synchronized pulsing of the flash unit is done via a LM1881-sync seperator integrated circuit which extracts the vertical sync signal from the composite video stream. The flashlight we are using delivers a reasonable brightness and contrast - much better than a strobe light in the same power range - but it needs approx. 0,8 seconds for completely recharging its capacitor. This problem was solved by adding a delay-and-shift circuit using a dual retriggerable monostable multivibrator (HEF4538, timing: cp. figure 7).
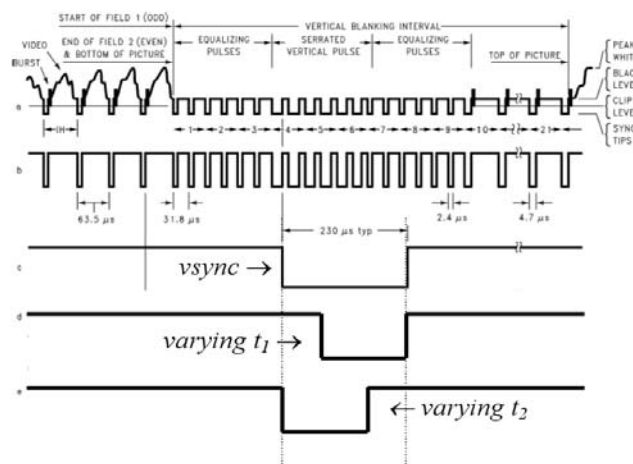

FIGURE 1. (a) Composite Video; (b) Composite Sync; (c) Vertical Output Pulse;

Fig. 7: Video-Timing for Flash-Sync (c: vsync, d,e: output), [19]

Two trimpots allow the user to adjust shift (delay) and pulse

width of the signal in a wide range of approx. 1,5 s. When connecting the projector with this sync circuit to the system, about every 20th frame will be flashed. To avoid the need of another synchronization the lighted frame is found by summing up all gray values in one diagonal and by rating the result using a threshold.

Fig. 8: Scan Unit mounted on the robot ALBERT

Unfortunately we cannot show any 3D scan results using the new miniaturized unit, because the manufacturing of the new calibration rig was not finished in time. But what we can say about the system is that it should deliver equally good results, because the images recorded using the new projector are of comparable quality. (figure 9)

Fig. 9: Screenshot recorded using the new Scan Head

2532

C. Software Environment

The algorithms as described above were implemented in C++ using Windows 2000 and the development environment Microsoft Visual Studio 6. The frame grabber was interfaced utilizing the VfW (Video for Windows) interface. The graphical user interface was realized with a Win32 API dialog. For online visualization a VTK 3D render window was used [14] but has currently been changed towards OpenGL to allow faster refresh rates in the process.

IV. RESULTS

A. Examples and Screenshots

In the first test runs of the beamer system we tried to scan a variety of different objects: a textured box (a packet of cigarettes, a small goose figurine made of plaster, pieces of chalk, a computer mouse). Here we want to show screenshots of the program's user interface and of the resulting 3D point clouds. (figure 10-12)
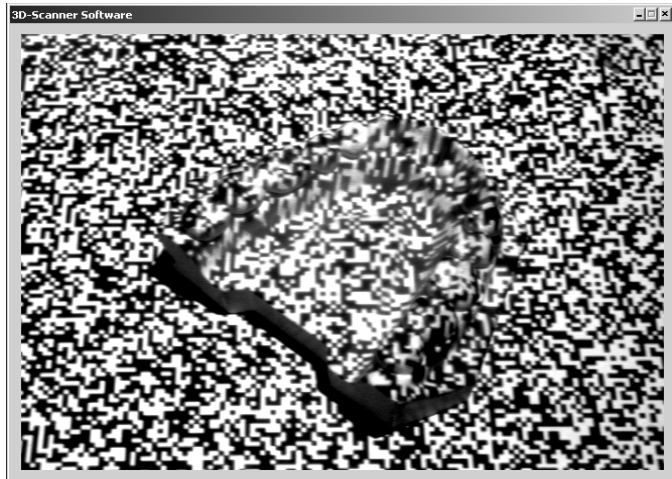

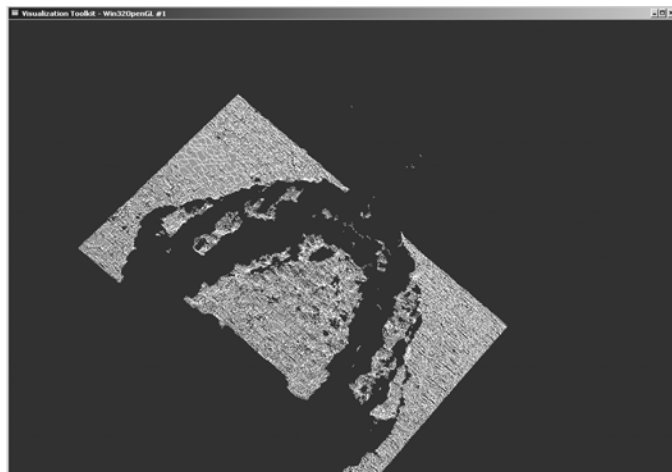Fig. 10. Teeth plaster cast with beamer-projected speckle pattern.

Fig. 11. Resulting sample scan (point cloud) of the plaster cast.

The white noise pattern can be recognized easily. We have chosen a relatively rough resolution of 320x240 in order to benefit from over-sampling in the correlation procedure. The resolution of the camera is 640x480.
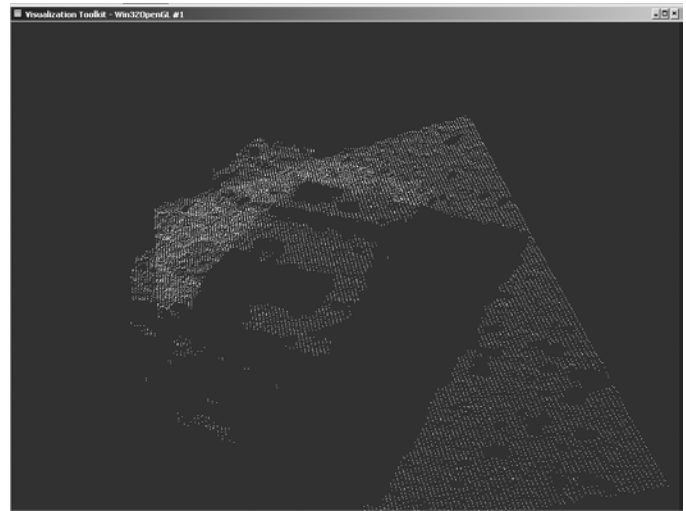

Fig. 12. Resulting scan (point cloud) of the packet of cigarettes.

B. Considerations on Accuracy

Due to the fact that the current implementation had only just been finished when this paper was written, we have not yet managed to perform intensive test runs. Anyway the calibration procedures deliver average errors of approx. 0.03 mm for the camera calibration and approx. 0.08 mm for the projector calibration. Of course this has to be put in relation to the chosen focal length of the camera (zoom, approx. 12 mm, CCD-to-object distance approx. 250 mm, 2/3'' CCD). Another preliminary test was to calculate a regression plane through a point cloud of the scan of a plane, delivering a maximal deviation of approx. 0.5 mm and an average deviation of approx. 0.2 mm.

V. CONCLUSION AND FUTURE WORK

In this paper we have described an effective and efficient approach to 3D shape acquisition based on a SSD correlation approach. The algorithms outlined above have been implemented in a test application and their functioning has been proven. The test-bed hardware consisting of a video projector and a digital camera on a frame assembly is easily scalable in geometry, resolution and focal distance as we have shown building the wrist mounted mini projector.

The next step will be to do a theoretical consideration of the system's errors and the achievable accuracy. In this context the SSD correlation algorithm, which is the key to resolution and accuracy in our correlation approach, has to be adapted and improved further upon for application in this special white noise pattern based system.

At the moment the camera images are preprocessed by simply spreading the gray value ranges in tiles of a fixed size, in order to obtain a normalized form as a basis for finding correspondences on surfaces of different colors. This procedure is still to be improved, primarily by incorporating it into the resampling procedure of the correlation algorithm.

The described ICP variant is already implemented and works fine with test data sets. Two important optimizations, namely the implementation of a kd-tree data struture and a calculation to find the overlapping regions of two datasets are still to be done.

When the calibration of the new miniaturized scan head is finished, we want to integrate this sensor in the grasp planning system of ALBERT expecting an improvement of the acquired near field 3D data in comparison to the global scene data delivered by ALBERT's stereo vision head. [17,18,20]

REFERENCES

[1] H. Hoppe, S. Däuber, C. Kübler, J. Raczkowsky, H. Wörn, "A New, Accurate and Easy to Implement Camera and Video Projector Model", in *Medicine Meets Virtual Reality, Conf. 02/10 (MMVR)*, Newport Beach, January 2002, pp.204-206

[2] A. W. Gruen, "Adaptive Least Squares Correlation: A Powerful Image Matching Technique", in *South African Journal of Photogrammetry, Remote Sensing and Cartography, Vol. 14, No. 3*, 1985, pp. 175-187

[3] N. D'Apuzzo, "Modeling Human Faces with Multi-Image Photogrammetry", in *Three-Dimensional Image Capture and Applications V. Proc. of SPIE*, San Jose, California, 2002, Vol. 4661

[4] T. Luhmann, *Nahbereichsphotogrammetrie: Grundlagen, Methoden und Anwendungen.* Wichmann, Heidelberg / Germany, 2000 (german)

[5] S. Rusinkiewicz, O. Hall-Holt, M. Levoy, "Real-Time 3D Model Acquisition", in *Siggraph 2002*, San Antonio / Texas / USA, July 2002

[6] S. Rusinkiewicz, M. Levoy, "Efficient Variants of the ICP Algorithm", in *Proc. of the 3. Int. Conf. on 3D Digital Imaging and Modeling (3DIM)*, Quebec / Canada, 2001

[7] P. J. Besl, N. D. McKay, "A Method for Registration of 3-D Shapes", in *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, No. 2,* February 1992, pp. 239-255

[8] F. Devernay, O. Bantiche, E. Coste-Maniere, "Structured light on dynamic scenes using standard stereoscopy algorithms", in *Institut National de Recherche en Informatique et en Automatique (INRIA), Research Report 4477*, June 2002

[9] P. Vuylsteke, A. Oosterlinck, "Range Image Acquisition with a Single Binary-Encoded Light Pattern", in *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 12, No. 2,* February 1990, pp. 148-164

[10] M. Maruyama, S. Abe, "Range Sensing by Projecting Multi-Slits with Random Cuts", in *International Workshop on Industrial Applications of Machine Intelligence and Vision (MIV-89)*, Tokyo, April 10-12, 1989, pp. 163-168

[11] M. Proesmans, L. J. Van Gool, A. J. Oosterlinck, "Active Acquisition of 3D Shape for Moving Objects", in *Int. Conf. on Image Processing*, Lausanne, September 1996, pp. 647-650

[12] E. Mouaddib, J. Batlle, J. Salvi, "Recent Progress in Structured Light in order to solve the Correspondence Problem in Stereo Vision", in *Proc. of the 1997 IEEE Int. Conf. on Robotics and Automation*, Albuquerque, New Mexico, April 1997, pp. 130-136

[13] E. Trucco, A. Verri: *Introductory Techniques for 3-D Computer Vision.* Prentice Hall, USA, 1998

[14] vtk, visualization toolkit, Kitware Inc., free download via http://www.vtk.org

[15] P. Azad, T. Gockel and R. Dillmann, "3D Shape Acquisition using a combined SSD and Least Squares Correlation Approach", *in: Proc. of the IEEE Int. Conf. on Information Technology (ITCC04)*, April 2004, Las Vegas, USA

[16] T. Gockel, P. Azad and R. Dillmann, "Calibration Issues for Projector-based 3D-Scanning", *in: Proc. of the IEEE Int. Conf. on Shape Modeling and Applications (SMI04)*, June 2004, Genova, Italy

[17] A. T. Miller, S. Knoop, H. I. Christensen, P.K. Allen, "Automatic Grasp Planning Using Shape Primitives", *in: Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA03),* Sept. 2003, Taipeh, Taiwan

[18] M. Ehrenmann, R. Becher, B. Giesler, R. Zöllner, O. Rogalla and R. Dillmann, „Interaction with Robot Assistants: Commanding ALBERT", *in: Proc. of the IEEE Int. Conf. on Intelligent Robots and Systems (IROS 02)*, Lausanne, Switzerland

[19] LM1881-Sync-Separator,Datasheet: *http://www.schaltungen.de /datenblatt/ datenblatt/LM/LM1881.pdf*

[20] A.T. Miller and P.K. Allen, „GraspIt!: a versatile simulator for grasping analysis", in: *Proc. of the ASME Dynamic Systems and Control Division, Vol. 2, pp. 1251 - 1258,* Orlando, FL, 2000.