DEVELOPMENT OF APPROACHES FOR 3-D HUMAN MOTION BEHAVIOUR ANALYSIS BASED ON RANGE IMAGING DATA

Patrick WESTFELD
Technische Universität Dresden, Institute of Photogrammetry and Remote Sensing, Germany
patrick.westfeld@tu-dresden.de

KEY WORDS: Integrated 3-D least-squares matching, range imaging, human motion behaviour analysis

ABSTRACT
Range imaging sensors generate intensity and range images simultaneously. The automatic evaluation of range image sequences requires the extension of 2-D image analysis procedures to 3-D. In this paper, a new approach for 3-D least-squares matching based on 2.5-D range imaging data is proposed. Thereby, 3-D motion vectors from intensity and depth image data are determined. It is possible to increase accuracy, stability and reliability of the least-squares method by combining two kinds of observations in one model and link them through geometric relations. By the example of an application in the field of pedagogics and educational science, approaches for the adaptation, further and new development of image analysis procedures to range imaging data are proposed and first results of the measurement of body distance and orientation as well as head inclination are presented.

1. INTRODUCTION AND MOTIVATION

In the field of pedagogics and educational science, methods for measurement and investigation of human behaviour have slowly evolved from simple video graphic recording, followed by interactive evaluation, to machine supported data acquisition and automatic or rather automated analysis of the collected data (Faraway, 2002; Schiano et al., 2000). The investigation of the interaction process, i.e. the communication among the studied individuals (so called »involvement«), constitutes a central point within behaviour research. Existing procedures for measuring involvement consider the channel of facial expression, eye movement and head inclination (Ekman & Friesen (1978, 1984)). The description of motion behaviour has also been realized by recording the entire body movement and spatial spacer sizes (Frey, 1984; Tronick, 1977). Furthermore, Graham et al. (1975) and von Salisch (1991) assumed that the nonverbal channels mentioned above provide the highest information content for educational surveys.

Video sequences can be evaluated interactively or (semi-) automatically. An interactive evaluation uses the context understanding of the operator. However, the associated subjectivity as well as a spatial and temporal generalisation of the recorded behaviour data is unfavourable. With regard to efficiency and objectivity, thus methods of photogrammetry and image analysis are particularly suitable to evaluate fine-dissolved involvement process data. That is a new and challenging field of application for photogrammetry. So far, photogrammetric image analysis has been primarily concerned with 3-D data acquisition for technical/industrial applications. Concerning tasks in human behaviour analysis, an increase in efficiency of the evaluation procedures as well as a significant increase in the spatial and temporal dissolution degree is to be achieved by a target-oriented adaptation and an enhancement of photogrammetric procedures.

Range imaging sensors (3-D cameras, distance-measuring cameras) allow the simultaneous acquisition of intensity and range images of – in principal – any object scene. Additional to the gray value information, the distances to the corresponding object points are measured for every pixel (cf.

chapter 2). As a result, a spatio-temporal resolved representation of the object space is given. In the field of range imaging sensor technology, range cameras are currently available with a sensor size of up to 25,000 pixels and a frame rate up to 50 fps. Advantages of this new 3-D mapping technology are the generation of 3-D data on a discreet raster without a stereocompilation, the recording of motion sequences as well as the marginal dimension and acquisition costs. Disadvantageously are the limited range, the small spatial resolution and the absolute accuracy in the range of a few centimetres. Possible applications for a range imaging sensor could be in the field of Human-Computer-Interaction (HCI), medical science, entertainment electronics, automotive engineering or pedagogics and educational science (cf. chapter 4).

The automatic evaluation of range image sequences requires the extension of 2-D image analysis procedures to 3-D. In the simplest case, processing techniques from greyscale image analysis can be applied directly to depth images. Regarding the nature of such data, suitable amendments and extensions of existing algorithms have to be made. In addition, it is necessary to solve the problem of data fusion during simultaneous evaluation of intensity and depth images.

In this article, an integrated 3-D least-squares matching (LSM) – an extended approach based on range imaging data – will be presented (cf. chapter 3). Thereby, 3-D motion vectors from intensity and depth image data are determined. It is possible to increase accuracy, stability and reliability of the least-squares method by combining two kinds of observations in one model and link them through geometric relations. The weighting of different kinds of observations (intensity and range) is made by the a priori accuracy characteristic of the data or by iterative weight estimation. Furthermore it is essential to develop a suitable parameterization of the geometric transformations. Besides an increase in accuracy and reliability, a differentiation between temporal object deformations and depth variations is possible.

The proposed integrated 3-D LSM provides a basis for the development of motion analysis methods based on range imaging data. The present paper gives an overview of new and adapted algorithm for 3-D human motion behaviour analysis (cf. chapter 4).

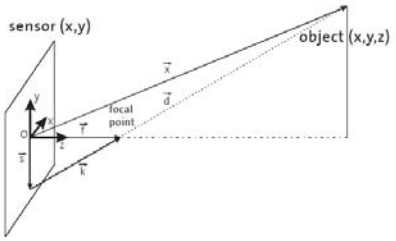
2. RANGE IMAGING SENSORS

Currently, geodetic instruments (tacheometer), multi-ocular photogrammetric recording configurations and terrestrial laser scanners are in operational use for 3-D object mapping. Tacheometers and laser scanners provide high accurate 3-D data in object space. However, they are less suitable for the registration of motion sequences because of the discrete and sequential nature of their measurements. Conventional photogrammetric procedures generate – depending on the sensors used – object space maps with high spatial and temporal resolution. The disadvantage is the recording configuration of at least two cameras, synchronized and referenced to each other, as well as the data processing, which is highly complex due to spatiotemporal feature matching.

Range imaging sensors offer an interesting monocular alternative for 3-D data acquisition. At this time, range cameras with a sensor size of up to 25,000 pixels and a frame rate up to 50 fps (e.g. *SwissRanger 3000* (CSEM, 2006); *PMD [vision] 19k* (PMD, 2007); Fig. 1a.) are available.

Figure 1: a. Left: *SwissRanger 3000* (CSEM, 2006). Right: *PMD [vision] 19k* (PMD, 2007)
b. Left: NIR intensity image. Right: Colour coded range image

The use of modulation techniques allow the acquisition of depth images synchronously to intensity images (Fig. 1b) by analyzing the intensity of emitted near infra-red light (NIR), which is reflected by the object. Additional to the known 2-D sensor position (x, y), each pixel of these images contains the information of the distance z to the corresponding object point. By continuous volume-like scanning of the object space, temporarily high-resolution 3-D data (**X**) on a discrete raster can be obtained after transformation of the 2.5-D data into a Cartesian coordinate system (Fig. 2).



$$\mathbf{X} = \mathbf{d} + \mathbf{s} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} + \begin{pmatrix} -x \\ -y \\ cc \end{pmatrix} \frac{z}{\sqrt{x^2 + y^2 + cc^2}} \tag{1}$$

with x,y: Image coordinates
cc: Focal length
z: Measured distance

Figure 2: Relation between image and object space (Kahlmann, 2006)

By avoiding stereoscopic matching, 3-D motion analysis becomes much simpler. For interdisciplinary experimental configurations the small dimension and initial costs are likewise favourable. At present, the limited range, the small spatial resolution as well as the accuracy in the range of a few centimetres are (still) cumbersomely. Together with suitable procedures for the determination of structured 3-D motion information from 3-D image sequences, range imaging sensors represent a versatile tool for data acquisition in 3-D motion analysis.

3. INTEGRATED 3-D LEAST-SQUARES MATCHING

In the following chapter, an extended least-squares matching approach based on 3-D camera image sequences is proposed. With respect to the range imaging data structure, a closed solution for intensity and range observations was developed. Besides an increase in accuracy and reliability, a differentiation between temporal object deformations and depth variations is possible.

3.1 Geometric And Radiometric Model

The basic least-squares matching approach formulates the gray value relation between two or more corresponding image patches as non-linear observation equations (Ackermann, 1984; Förstner, 1984; Grün, 1985). The goal is to determine the parameters of a 2-D affine transformation and – if necessary – a radiometric correction term as well. Commonly used for spatial and/or temporal matching tasks (e.g. conventional aero triangulation, DSM generation or motion analysis applications), LSM represents a multifunctional instrument for image analysis.

With respect to range imaging sensors, it appears desirable to adapt the primal LSM algorithm to the 2.5-D data structure. A trivial solution could be the formulation as a multi-channel adjustment with at most restrictions between intensity and range images. From a statistical point of view this approach is not optimal. The article presents a closed solution for intensity and range observations. Based on the 2.5-D least-squares adjustment of TIN structures (Maas, 2000), both kinds of observations are integrated in one model to increase accuracy and reliability.

The introduction of NIR intensity observations happens as usual. Template patch $g_1(x, y)$ and search patch $g_2(x, y)$ provide gray value observations for the adjustment. LSM compares the image contents and tries to minimize the sum of the gray value differences between those windows. If the geometric transformation between the two patches can be modelled by two relative translations in row and column direction (a_0, b_0), two scale adjustments in both coordinate directions (a_1, b_2) as well as two rotation parameters (a_2, b_1), the observation equation for each pixel becomes

$$\Delta g(x,y) = r_0 + r_1\, g_2\,(x',y') - g_1(x,y) \tag{2}$$

At this, the geometric affine transformation is

$$x' = a_0 + a_1 x + a_2 y \quad \text{and} \quad y' = b_0 + b_1 x + b_2 y \tag{3}$$

r_0 as well as r_1 – the parameters of a linear radiometric function – model radiometric differences between the gray values of template and search patch.

To consider the 3-D nature of the range measurements, one additional range observation per pixel within the patches is adapted into the basic least-squares matching model. The range variations between template and search window could be formulated as a linear function. Thus, it is possible to estimate the 1-D depth shift factor d_0 as well as the depth scale factor d_1 simultaneously. Thereby, the geometric relation between (x, y) and (x', y') is identical to equation (3).

$$\Delta z(x,y) = d_0 + d_1 z_2(x',y') - z_1(x,y) \tag{4}$$

3.2 Parameterization of Integrated 3-D LSM and Weighting of the Observations

As mentioned above, the geometric 2-D affine transformation between two small and even patches is done by six parameters – two shifts and scales in x and y, one rotation, one shear. Due to the new introduced depth shift as well as the depth scale, correlations between the basic LSM parameters and the new one may be expected. Therefore, a suitable parameterization should be developed.

As shown in Fig. 3 (left), a depth variation of the point of interest causes a scale adjustment in the 2-D affine parameter set. That adjustment could be substitute by additional range observation equations (cf. equation (4)) for every pixel within the patches. Thus it is possible to express a shift in depth direction directly by an observation and increase therewith the stability and reliability of the least-squares approach. Consequently it is necessary to reduce the set of parameters by a_1 and b_2 in order to avoid an over-parameterization of the transformation model.

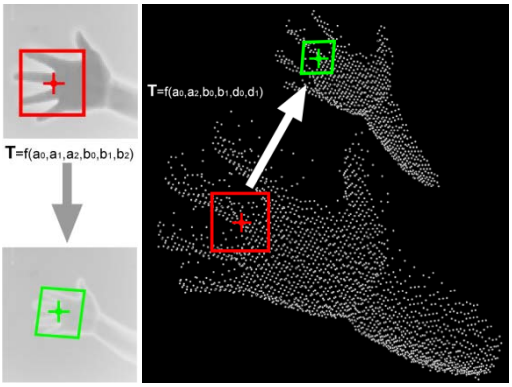


Figure 3: Relation between basic 2-D LSM (left) and integrated 3-D LSM (right); not to scale.

The weighting of different kinds of observations (intensity and range) is made by the a priori accuracy characteristic of the data or (in future) by iterative weight estimation (estimation of variance components).

First accuracy trials yielded the following results:

- The accuracy of intensity measurement mainly depends on the object distance (degradation of energy of the emitted NIR wave front with increasing distance), the influence of back-ground radiation and the charge-to-voltage converter for each single pixel.
- The accuracy of distance measurement mainly depends on the object distance and the inte-gration time (signal-to-noise ratio).

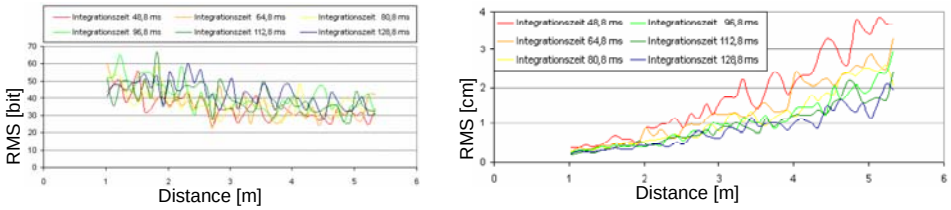


Figure 4: Left: RMS deviation for intensity measurements (scale unit in 16 bit) increases with increasing integration time and decreasing object distance. Right: RMS deviation for distance measurements (scale unit in cm) increases with decreasing integration time and increasing object distance.

According to the average noise of intensity and range measurements (cf. Fig. 4), the a priori weight ratio is set up to 1 (range observation) to 10 (intensity observation) in the first instance.

3.3 Experimental Evaluation

In this chapter, the main focus lies on the evaluation of the utilisability of the proposed 3-D LSM algorithm as well as the accuracy potential. First a simplified assumption was met, as a scale adjustment within the range patches is not expected ($d_1 := 1$; cf. equation (4)). According to (Baltsavias, 1991), the radiometric correction terms were determined prior the actual least-squares adjustment, to yield a robust and accurate solution for the remaining geometric parameters. Due to the introduction of the new parameter for depth shift, it was necessary to analyze the significance of this parameter for the whole model. This significance test (one-dimensional Student test with a quantile for a 95% confidence coefficient) allows the detection of an over-parameterization and supports the assumptions made in chapter 3.2.

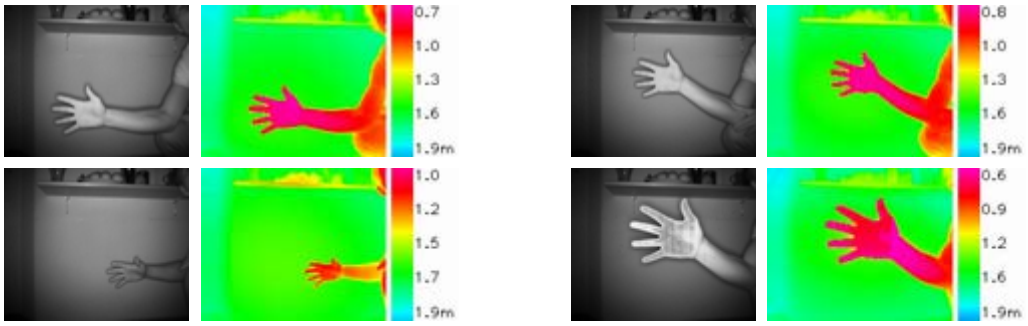


Figure 5: Experimental series detail (Left: Template images. Right: Search images.)

Table 1: A posteriori standard deviations $\hat{\sigma}_i$ of the parameters and mean corrections $\overline{v}_{GV/z}$ of gray value and range observations for some significant range camera test series.

$\hat{\sigma}_i$ Trial	a_0 [Pixel]	a_1 [Pixel]	a_2 [Pixel]	b_0 [Pixel]	b_1 [Pixel]	b_2 [Pixel]	d_0 [mm]	d_1 [mm]	$\overline{v}_{GV}$ [16 bit]	$\overline{v}_z$ [16 bit]
T1_1	0,061	0,042	0,040	0,060	0,020	0,016	-	-	585,25	-
T1_2	0,014	0,010	0,009	0,014	0,004	0,003	6,096	*fix*	581,07	170,85
T2_1	0,161	0,076	0,044	0,084	0,030	0,030	-	-	952,17	-
T2_2	0,035	0,016	0,009	0,018	0,006	0,006	8,788	*fix*	861,24	283,55
T3_1	0,155	0,069	0,070	0,098	0,040	0,048	-	-	722,70	-
T3_2	0,083	0,037	0,038	0,052	0,021	0,026	16.407	*fix*	690,50	1785,15
T4_1	0,270	*non-sign.*	0,081	*non-sign.*	0,046	0,057	-	-	1356,65	-
T4_2	0,095	0,040	0,024	0,061	0,021	0,017	17,648	*fix*	1319,42	1042,67
T5_1	0,132	*non-sign.*	*non-sign.*	0,105	*non-sign.*	0,037	-	-	2054,21	-
T5_2	0,040	0,026	*non-sign.*	0,032	0,012	0,012	17,67	*fix*	1617,70	618,55

The results of several performed experimental series (cf. Fig. 5) are shown in Table 1. For every trial *Ti*, a basic 2-D LSM (*Ti_1*) and an integrated 3-D LSM (*Ti_2*) were performed. The a posteriori standard deviations of the computed parameters as well as the mean corrections of the observations are set in contrast with each other in the above mentioned Table 1.

As shown in Table 1, an increase in accuracy by a factor *2* to *5* could be determined for the geometric affine parameters of *Ti_2* in contrast to *Ti_1*. Furthermore, the mean corrections were computed, in order to prove, that the increase in accuracy is not solely due to doubling the number of observations. A minimization of the corrections $\overline{v}_z$ in comparison with $\overline{v}_{GV}$ is observable. It can be assumed, that the use of the range observations in combination with the additional depth shift parameter improve the geometric modelling of 2.5-D range imaging data.

4. PHOTOGRAMMETRIC 3-D HUMAN MOTION BEHAVIOUR ANALYSIS

4.1 General Aspects

Relevant for the registration of the involvement within the field of behaviour research are:

- Interpersonal distance
- View behaviour and head inclination
- Body orientation
- Facial gestures

Due to the limited sensor format (cf. chapter 2), view behaviour and facial gestures of the probands cannot be evaluated by a 3-D camera. Therefore, the scope of this section is limited to the identification of approaches and the presentation of first results for the adaptation, further and new development of image analysis procedures of body distance and orientation as well as head inclination.

The above introduced integrated 3-D LSM provides a basis for the development of motion analysis methods based on range imaging data and is used for the following temporal matching tasks. Furthermore, in the first instance, all necessary segmentations of the probands were realized in a semiautomatic way (cf. Figure 6). This means, that the segmentation was simply accomplished due to range information. The spatial separation of the probands to each other was done interactively.

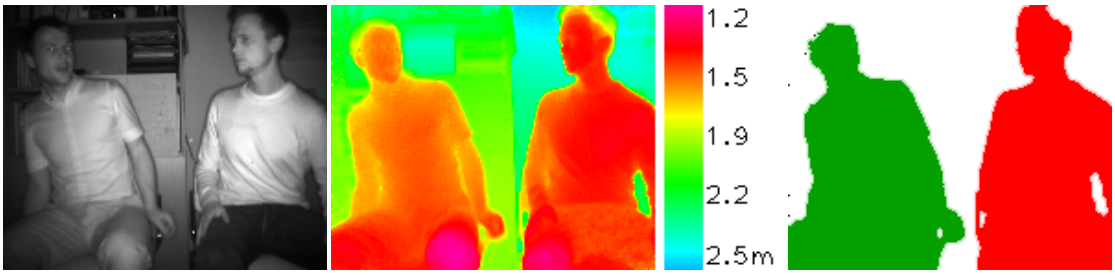


Figure 6: Left: Intensity image. Middle: Colour coded range image. Right: Image segmentation.

4.2 Interpersonal Distance

Figure 7 (down right) shows colour coded depth images of the *SwissRanger 3000*. The outlines of the persons are extractable by segmentation procedures (cf. chapter 4.1). From these outlines the interpersonal distance can be determined; if necessary after a reduction to the centroid.

The trajectories of two probands during an interaction are shown in Fig. 7 (left). A tracking of the centroid – determined in the first image – through all images of the sequence was realized with the proposed integrated 3-D LSM.

Due to the 3-D nature of the measurements, the distance is given in object space and no further transformation is needed. The evaluation of the image sequences results in a temporal variation of the interpersonal distance with a temporal resolution of up to 50 fps.

The a posteriori standard deviation of the translation parameters in and are in the range of 0.01-0.02 pixel; the depth shift parameter is in the range of 2-3 mm.

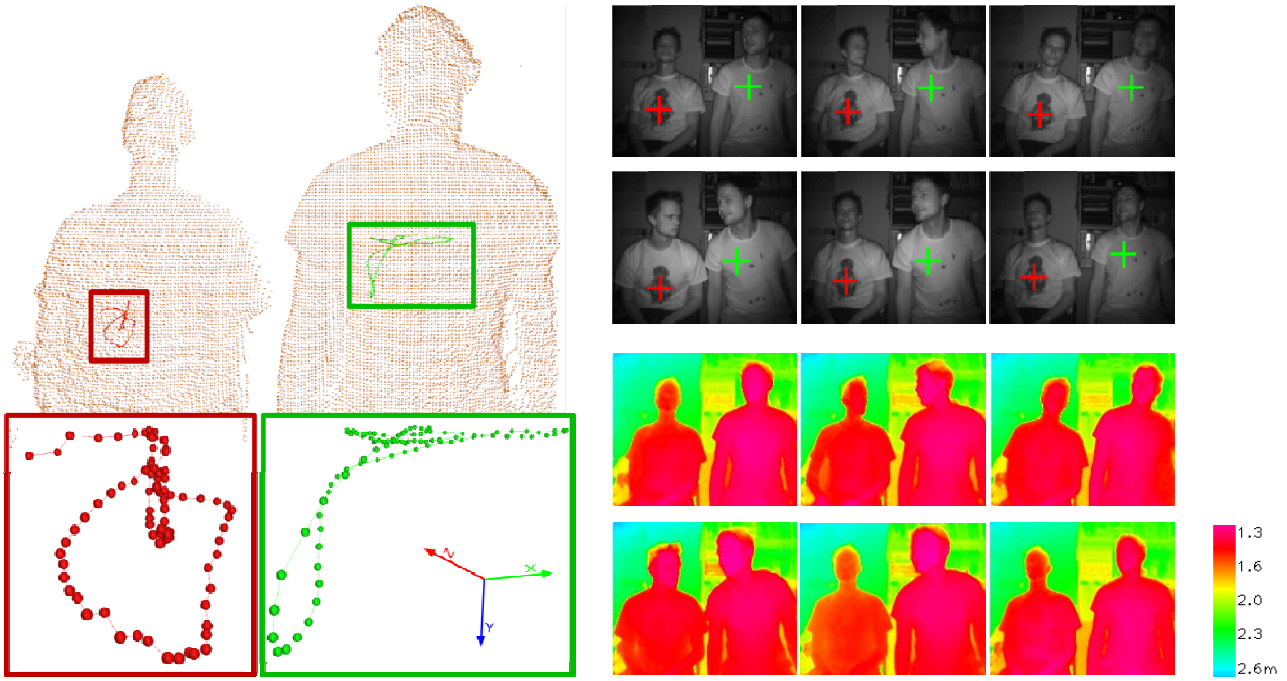


Figure 7: Left: 3-D point clouds of the probands including 3-D centroid trajectories computed by integrated 3-D LSM. Right: NIR intensity images with tagged feature positions and colour coded range images.

4.3 Body Orientation

Besides getting information on interpersonal distance, information about the relative body orientation between two persons can be gained. After a segmentation and centroid reduction, the determination of the main inertia axes (principal components of the respective data sets) is possible.

Principal Component Analysis (PCA) is a multivariate statistic method with the objectives of dimension reduction and pattern identification (e.g. Fukunaga, 1990). In the present case, PCA is used for the determination of the main component axes only, whereby the eigenvectors (direction of the axes) can be computed by the covariance matrix. A quantity for the variation of body orientation results from the computation of the angles between the three several principal components (PC) through the whole image sequence. Figure 8 shows a first example. By segmentation of the probands, a principal component system is available for both persons and allows determining the relative body orientation to each other (cf. Fig. 9).

Problematic – but not examined until now – are the dependence of PCA on outliers because of measurement uncertainty and extremity motions without body motion. Furthermore, an acquisition of whole bodies (front and rear side) is not possible using one 3-D camera only. Therewith, an equal distribution of the points within two frames is possible, even though body movement (cf. PCA applied on a sphere).

To avoid this circumstance, significant features (e.g. points on the left and right shoulder) could be tracked with depth and texture based matching algorithm (integrated 3-D LSM) through an image sequence in three dimensions. After spanning a vector between those points, the vector angle could be computed in the same manner as the relative orientation between one principal component at two epochs. Thus, the calculation of the variation in the body orientation is also feasible. Another

possibility is to perform a discrete sub-segmentation of the upper part of the body as well as the arms.

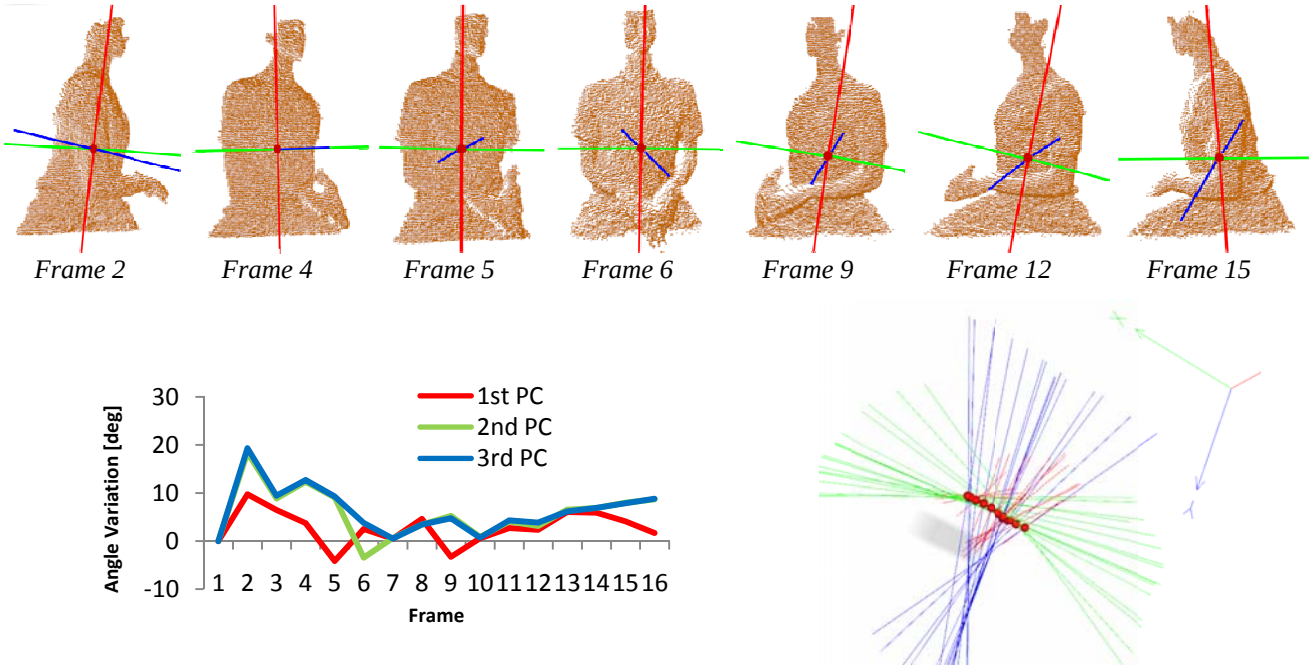


Figure 8: Top: Determination of body orientations by PCA (1st PC red coloured; 2nd PC green coloured; 3rd PC blue coloured). Down: Variations in body orientation; angle variations of the several PCs within one image sequence.

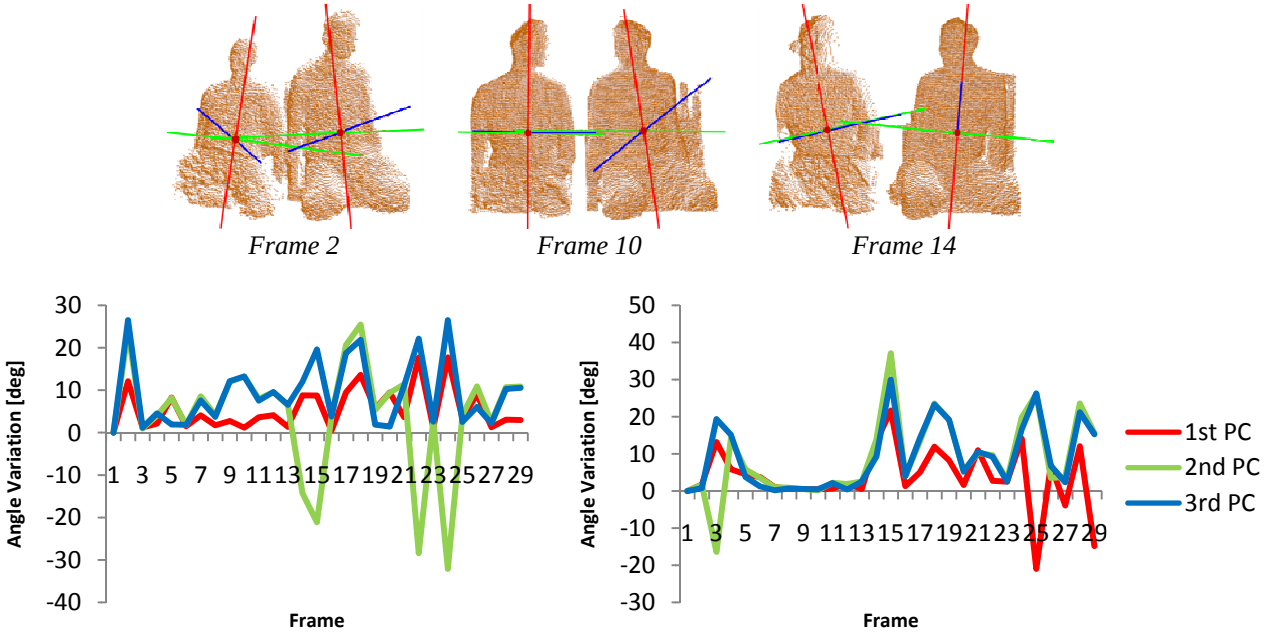
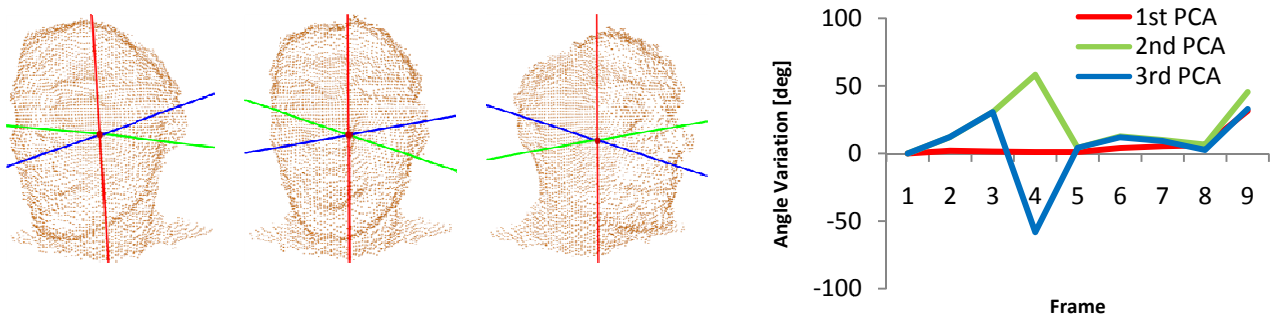


Figure 9: Top: Determination of relative body orientations between two probands. Down: Angle variations in body orientation of proband #1 (left) and proband #2 (right).

4.4 Head Inclination

According to the remarks of chapter 4.3, the subject's head movement is feasible after making another sub-segmentation. Graphical results are shown in Figure 9.



Frame 2 *Frame 5* *Frame 8*

Figure 10: Determination of head inclination.

Also in this case, the reliability of the approach is constricted because of equal point distributions (cf. chapter 4.3) but not examined until now.

5. CONCLUSIONS AND OUTLOOK

In this article, a new least-squares approach and photogrammetric approaches for 3-D human motion behaviour analysis – both based on range imaging data – were proposed.

Due to the development of range imaging sensors, it seems desirable to extend 2-D matching procedures to 3-D. Thus it is possible to differentiate between real depth variations and deformations within a plane. It could be shown, that an increase in accuracy, stability and reliability for least-squares matching can be reached by a suitable model adaption with respect to the 2.5-D range imaging data nature. The weighting of different kinds of observations as well as a suitable parameterization were discussed and first results were shown. Furthermore, an interdisciplinary application in the field of human behaviour analysis was introduced. Using a range imaging sensor in conjunction with automatic analysis methods allow the acquisition of fine-dissolved involvement process data. The present article discloses first efforts for the measurement and (semi-) automatic analysis of interpersonal distance, body orientation and head inclination.

Future work will concentrate on an improved least-squares weighting of different kinds of observations by means of estimation of variance components. Furthermore, the exterior accuracy potential of the presented integrated 3-D least-squares matching is to be determined. Because of the noise characteristics of the 3-D camera, a combination of object tracking and surface smoothing is to be analysed. Beyond that, further aims are the development of automatic (sub-) segmentation strategies of the human body, the improvement of PCA applied on range imaging point clouds and accuracy tests of the proposed motion analysis methods.

6. ACKNOWLEDGMENTS

I would like to thank my cooperating partners Prof. Dr. phil. habil. L.-M. Alisch, M.A. Rico Hermkes and M.A. Uwe Altmann (Technische Universität Dresden, Faculty of Education, {lutz-michael.alisch, rico.hermkes, uwe.altmann}@mailbox.tu-dresden.de) for helpful, supportive as well as enjoyable suggestions and discussions.

REFERENCES

1. Ackermann, F., 1984: High Precision Digital Image Correlation. Proceedings of the 39th Photogrammetric Week, volume 9. Schriftenreihe der Universität Stuttgart. Stuttgart.
2. Baltsavias, E. P., 1991: Multiphoto Geometrically Constrained Matching. Ph. D. Dissertation. Report No. 49. Institute of Geodesy and Photogrammetry. ETH Zurich. Switzerland.
3. CSEM, 2006: SwissRanger SR-3000. Miniature 3D time-of-flight range camera. Centre Suisse d'Electronique et de Microtechnique. Zurich. Switzerland. http://www.swissranger.ch/. (May 2006).
4. Ekman, P. & Friesen, W.V., 1984: Manual EMFACS-7.
5. Faraway, J., 2002: The average smile-modeling continuous shape change for facial animation. Technical report # 386. Department of Statistics. University of Michigan. USA. www.stat.lsa.umich.edu/ ~faraway/papers/smilepub.pdf.
6. Förstner, W., 1984: Quality assessment of object location and point transfer using digital image correlation techniques. IAPRS Vol. 25, Part A3a, pp. 197-217.
7. Frey, S., 1984: Die nonverbale Kommunikation. SEL- Stiftung für technische und wirtschaftliche Kommunikationsforschung im Stiftungsverband für die Deutsche Wissenschaft.
8. Fukunaga, K., 1990: Introduction to Statistical Pattern Recognition. 2nd Edition. Elsevier. ISBN 0122698517.
9. Graham, J.A.; Ricci Bitti, P. & Argyle, M., 1975: A cross-cultural study of the communication of emotion by facial and gestural cues. Journal of Human Movement Studies. 178-182.
10. Grün, A., 1985: Adaptive Least Squares Correlation - A Powerful Image Matching Technique. South African Journal of Photogrammetry, Remote Sensing and Cartography. Vol.14, No. 3.
11. Kahlmann, T., Remondino, F., Ingensand, H., 2006: Calibration for Increased Accuracy of the Range Imaging Camera SwissRanger. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVI, Part 5, pp. 136-141.
12. Maas, H.-G., 2000: Least-Squares Matching with Airborne Laserscanning Data in a TIN Structure International Archives of Photogrammetry and Remote Sensing. Vol. 33, Part 3A, pp. 548-555.
13. PMD, 2007: PMDTechnologies GmbH. Siegen. Germany. http://www.pmdtec.com/. (January 2007).
14. Salisch, M. von, 1991: Kinderfreundschaften. Emotionale Kommunikation im Konflikt. Hogrefe. Göttingen, Toronto, Zürich.
15. Schiano, D., Ehrlich, S., Rahardja, K., Sheridan, N., 2000: Measuring and modeling facial affect. In: Behavior Research Methods, Instruments & Computers. 32 (4), pp. 505-514.
16. Tronick, E.D., Als, H. & Brazelton, T.B., 1977: Mutuality in mother-infant interaction. Journal of Communication. 27. 74-79.