RANGE IMAGE SEQUENCE ANALYSIS BY 2.5-D LEAST SQUARES TRACKING WITH VARIANCE COMPONENT ESTIMATION AND ROBUST VARIANCE COVARIANCE MATRIX ESTIMATION

Patrick Westfeld[a,*] **and René Hempel**[b]

Technische Universität Dresden, D-01062 Dresden, Germany

[a]Institute of Photogrammetry and Remote Sensing (IPF), patrick.westfeld@tu-dresden.de, http://www.tu-dresden.de/ipf/photo
[b]Faculty of Education, rene.hempel@tu-dresden.de, http://tu-dresden.de/die_tu_dresden/fakultaeten/erzw/erzwiae/ewwm

KEY WORDS: Range Imaging, Least Squares Tracking, Variance Component Estimation, Robust Variance Covariance Matrix

ABSTRACT:

In this article, a range image sequence tracking approach is proposed, which combines 3-D camera intensity and range observations in an integrated geometric transformation model. Based on 2-D least squares matching, a closed solution for intensity and range observations has been developed. By combining complementary information, an increase in accuracy and reliability can be achieved. The weighting of the two different types of observations with a-priori unknown quality is performed by variance component estimation. To fulfill the requirements of robust variance covariance matrix estimation in statistical context, alternative approaches for variance covariance matrix calculation are proposed and evaluated. To verify its applicability, reliability and accuracy potential, the introduced 2.5-D least squares tracking technique has been evaluated by several series of experiments in the field of human motion and interaction measurement.

1 INTRODUCTION AND MOTIVATION

Conventional stereo-photogrammetric procedures generate, depending on the sensors used, object space maps with high spatiotemporal resolution. The main drawbacks are the recording configuration of at least two cameras, synchronized and oriented to each other, and the data processing, which is highly complex due to spatial and temporal feature matching.

Range imaging (RIM) cameras (3-D cameras) based on photonic mixer devices (PMD; Schwarte, 1997) or comparable principles offer an interesting monocular alternative for photogrammetric 3-D data acquisition. The use of modulation techniques and combined CCD/CMOS technology provides simultaneous gray value and distance measurements of the scene in each pixel of the sensor. With frame rates up to 50 Hz, 3-D cameras are well suited for motion capture in fields such as human or robot (inter-)action analysis.

Several approaches to (semi-)automatic RIM sequence analysis have been shown: Göktürk and Tomasi (2004) introduced a RIM head-tracking algorithm. In a training stage, a depth signature (representative signature for head location) is calculated by identifying the probands' heads on each frame interactively. In a tracking stage, the depth-signature of each frame is compared against the training signatures. The best match can be identified by a correlation metric and represents the location of the object of interest. Kahlmann and Ingensand (2006) described the usability of the RIM camera *SwissRanger SR-3000* for surveillance systems. Moving persons within an indoor scene could be detected by RIM thresholding and pixel clustering. Gesture recognition based on motion detection by double difference range images and 3-D shape matching with 3-D shape contexts has been presented by Holte and Moeslund (2007). Breuer et al. (2007) recognized hand movements (location and orientation) by principle component analysis (PCA) applied on RIM data. In the further course of analysis, they fitted an articulated model to reconstruct the hands. The centroid of a cluster represents the persons position for the corresponding frame. The implementation of the CONDENSATION algorithm (conditional density propagation) – first introduced by Isard and Blake (1998) and extended for tracking multiple objects in RIM sequences by Koller-Meier (2000) – into a RIM tracking process is described in Kahlmann et al. (2007).

The above reviewed RIM tracking approaches are based on basic image analysis functions (e.g. thresholding, segmentation, computation of point cloud centroid) or extended matching procedures using motion and measurement models (e.g. CONDENSATION algorithm, Kalman filtering) applied to the range data. In this article, a RIM sequence tracking approach (2.5-D least squares tracking; LST) is proposed, which combines RIM intensity and range observations in an integrated geometric transformation model. Based on 2-D least squares matching (LSM), a closed solution for intensity and range observations has been developed. In contrast to motion model techniques, intensity observations are also included into the least squares (LS) adjustment. By adding complementary information, an increase in accuracy and reliability can be expected.

2 SENSOR AND DATA

RIM sensors (Figure 1) allow the simultaneous acquisition of intensity and range images of – in principal – any scene (Figure 2). In the field of RIM sensor technology, 3-D cameras are currently available with a sensor size of up to 25,000 pixels and a frame rate of up to 50 Hz. Based on a phase-measuring time-of-flight (TOF) principle, the camera is able to measure distances for each pixel in addition to the gray value information (Oggier et al., 2004). As a result, a spatiotemporal resolved representation of the object space is given in the form of intensity images and range maps. The calculation of 3-D coordinates is performed on-chip. Image coordinates as well as range information are transformed into Cartesian coordinates using the relationship between image and object space as described in Kahlmann and Ingensand (2006). Several assumptions are implied, which have to be proven by suitable photogrammetric calibration techniques (Kahlmann et al., 2006; Westfeld, 2007a).

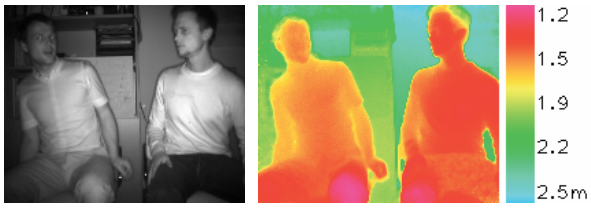
Advantages of this new 3-D mapping technology are the generation of 3-D data on a discreet raster without stereo compilation, the recording of motion sequences and the marginal dimension.

* Corresponding author.

Drawbacks are the limited range, the small spatial resolution and the absolute accuracy in the range of a few centimeters. Possible applications for RIM sensors could be in the field of human-computer-interaction (HCI; Du et al., 2005), robot vision (Gudmundsson, 2006), automotive engineering (Zywitza et al., 2005) or human motion analysis (Westfeld, 2007b; Hempel and Westfeld, 2008).

(a) (b)

Figure 1: Range imaging cameras. (a): *SwissRanger SR-3000* (url: http://www.swissragner.ch/, 2007). (b): *PMD [vision] 19k* (url: http://www.pmdtec.com/, 2007).



(a) (b)

Figure 2: Near-infrared intensity image (a) and colour coded range image (b).

The data used in this article were captured by the *SwissRanger SR-3000* (Mesa Imaging AG, Zurich, Switzerland; Figure 1a). It should be pointed out that alternative manufactures, like PMD-Technologies GmbH (Siegen, Germany; Figure 1b) or Canesta, Inc. (Sunnyvale, CA, USA), offer commercially available products, too. The modes of operations are nearly the same, except for the chip design: Mesa Imaging AG uses combined CCD/CMOS technology, PMDTechnologies and Canesta, Inc. just use CMOS. A detailed survey of optical range measurement and solid-state imaging sensing is given in Lange (2000).

3 RANGE IMAGE SEQUENCE ANALYSIS

Photogrammetric motion analysis is a well-established part of close-range photogrammetry and allows the extraction of geometric information from images with high precision and reliability. In this context, least squares matching is a common tool for the computation of motion vectors from image sequences.

3.1 State of the Art

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

3-D LSM: The basic 2-D LSM approach was extended to a 3-D algorithm working on voxel data and applied on flow tomography sequences by Maas et al. (1994). Accordingly, 3-D LSM works with 3-D volume data and voxels rather than 2-D images and pixels. The 3-D affine transformation has 12 parameters, and the observation equations have to be formulated using gray value gradients in three directions.

Least Squares Surface Matching: Based on a basic 2-D LSM approach, Maas (2000) computed correspondences between neighboring or crossing airborne laser scanning strips by formulating LSM on a TIN structure. Grün and Akca (2004) proposed a 3-D least squares surface matching algorithm (LS3D), which estimates the seven parameters of a 3-D similarity transformation between two or more 3-D surface patches by minimizing the Euclidean distances.

3.2 2.5-D Least Squares Tracking

2-D LSM can basically be applied for tracking surface patches in RIM data sequences by using the Cartesian coordinates only. The proposed 2.5-D LST (least squares tracking) algorithm uses the original intensity and range information simultaneously. Due to the 2.5-D nature of the surface patches, this is referred to as 2.5-D here.

Functional Model: Intensity observations are used in the same manner as in conventional LSM: Template patch $\mathbf{gv_1}$ and search patch $\mathbf{gv_2}$, taken from consecutive gray value images $\mathcal{I}_1$ and $\mathcal{I}_2$, provide gray value observations for the adjustment at each position (x, y) resp. (x', y'). The geometric and radiometric relations between those patches can be formulated as

$$\mathbf{gv_1}(x, y) - \mathbf{v_1}(x, y) = r_0 + r_1 \cdot \mathbf{gv_2}(x', y') \qquad (1)$$

Based on the same considerations, the relation between two patches $\mathbf{rv_1}$ and $\mathbf{rv_2}$ taken from range value images $\mathcal{R}_1$ and $\mathcal{R}_2$ become

$$\mathbf{rv_1}(x, y) - \mathbf{v_2}(x, y) = d_0 + d_1 \cdot \mathbf{rv_2}(x', y') \qquad (2)$$

The geometric affine transformation model for both types of observations, intensity and range, is given by

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

In Equation 1, r_0 and r_1 model brightness and contrast variations. In analogy to a radiometric gray value correction, range variations between template and search window can be formulated as a linear function, too. Thus, it is also possible to compute a 1-D depth shift d_0 and a depth scale factor d_1. Within a Gauss-Markov-Model (GMM), the parameters can be estimated by minimizing the sum of the squares of the error vectors $\mathbf{v_1}$ and $\mathbf{v_2}$.

Stochastic Model: The stochastic model describes the variances and covariances of the observations. In many cases, the setup of the stochastic structure of the observations (variance-covariance matrix; VC-matrix) is given by the a-priori definition of fixed weights. Information from the instrument manufacturer or from previous accuracy analyses provide the basis for the specification of the quality of the observations. Besides the parameters of the functional model, the standard error of unit weight can be estimated for the stochastic model only. Thus, the evaluation of the quality of the observations using the variance of the unit weight is limited to homogeneous observations.

2.5-D LST uses heterogeneous observations (intensity and range) and requires adjusted weights for each group of observations. In our work, the weights are computed by iterative variance component estimation (VCE; e.g. Schneider and Maas, 2007). This approach provides the following advantages:

- An automatic estimation of the variance components.

- An improvement of the adjustment results due to the utilization of the complete information content of the observations.
- An accuracy specification for each variance component and therewith for each group of measurement.

The weights of the observations are given by the quotient of the variance of the unit weight σ_0^2 and the variances of the observations σ_i^2. At this, σ_0 is a constant (in general $\sigma_0 = 1$) and σ_i are the variance components, which can be estimated. In the course of the VCE, the VC-matrix of the unknown parameters $V(\hat{\mathbf{x}})$ can be divided into additive components $V_i(\hat{\mathbf{x}}) = \sigma_i^2(\mathbf{A}'\mathbf{\Omega}^{-1}\mathbf{A})^{-1}$; each component i represents a group of observations. See Koch (2004) for further information for the computation.

Parametrization: Like in Baltsavias (1991), the radiometric gray value correction terms were determined prior to the actual LS adjustment. This approach yields a more robust solution for the remaining geometric parameters.

The geometric transformation between both pairs of patches $\mathbf{gv_i}$ and $\mathbf{rv_i}$ with $i = [1; 2]$ can be modeled by an affine transformation with two shifts a_0, b_0, two scales a_1, b_2 in row and column direction and two parameters a_2, b_1 for rotation and shear. The gray values as well as the range values remain unaffected by the shift and rotation parameters and are resampled only, according to their new (non-integer) image positions. The remaining scale parameters cause geometric variations in both search patches, but effect a change in range values, too. Consequently, the range offset d_0 depends on a_1,b_2 and can be integrated into the basic 2-D LSM approach, which allows a closed LST solution. Furthermore it was assumed that there is no depth scale variation in the range patches. Therefore, the depth scale parameter d_1 is set to 1 in the following considerations.

The relation between depth variations and – in a first instance – a consistent scale in row and column direction $\lambda := \frac{1}{2}(a_1 + b_2)$ is given by

$$\lambda = \frac{\mathbf{rv_1}(x, y)}{\mathbf{rv_2}(x', y')} \tag{4}$$

Due to scale-invariant range value differences, Equation 4 is expressed for the two center pixels (x^c, y^c) and (x'^c, y'^c). The range value adjustment of the center pixel becomes

$$\mathbf{rv_1}(x^c, y^c) = \lambda \cdot \mathbf{rv_2}(x'^c, y'^c) \tag{5}$$

Therewith, the range values of the remaining pixels in the neighborhood (x^n, y^n) and (x'^n, y'^n) can be formulated:

$$\begin{aligned}\mathbf{rv_1}(x^n, y^n) = \mathbf{rv_1}(x^c, y^c) \\ + \left[\mathbf{rv_2}(x'^n, y'^n) - \mathbf{rv_2}(x'^c, y'^c)\right]\end{aligned} \tag{6}$$

Substituting $\mathbf{rv_1}(x^c, y^c)$ in Equation 6 by Equation 5 yields a range value correction term according to scale variations for all pixels:

$$d_0 = \mathbf{rv_1}(x, y) - \mathbf{rv_2}(x', y') = \mathbf{rv_2}(x'^c, y'^c) \cdot (\lambda - 1) \tag{7}$$

Finally, the observation equation 2 can be expressed as

$$\begin{aligned}\mathbf{rv_1}(x, y) - \mathbf{v_2}(x, y) = \mathbf{rv_2}(x'^c, y'^c) \cdot \left(\frac{a_1 + b_2}{2} - 1\right) \\ + 1 \cdot \mathbf{rv_2}(x', y')\end{aligned} \tag{8}$$

Using this integrated model, all transformation parameters can be determined based on intensity and range observations. The GMM minimizes the sum of squares of the intensity and range value differences. The range offset d_0 in dependency of a_1, b_2 is considered in the GMM of observation vector.

4 MISSPECIFIED VC-MATRICES IN GENERALIZED MULTIPLE LINEAR REGRESSION MODELS AND THEIR CONSEQUENCES

In statistics, a random variable is called heteroscedastic (HS), if at least two different observations do not have the same variance (Greene, 2007). Estimating the variances of each observation is impossible due to the lack of redundancy and is not desirable from a geodetic point of view. As a result, a HS pattern has to be introduced in the form of an assumed VC-matrix of a vector valued random variable (in general by giving equal weights to each vector component). This assumption may cause inaccurate parameter estimations and invalid statistical hypotheses tests. A (partial) correction for heteroscedasticity can be achieved by the application of a weighted LS estimation method.

In this article, the weighting of the two different types of observations with a-priori unknown quality is firstly performed by VCE (Section 3). However, the variances of the two types of observations estimated by VCE may not correspond to the true variances of each individual observation. To fulfill the requirements of a robust VC-matrix estimation in statistical context, alternative approaches for VC-matrix calculation are proposed and evaluated, which yield estimators that are consistent for the true VC-matrix. These include the heteroscedasticity consistent estimator (HC; White, 1980) disclaiming any autocorrelation within the disturbances as well as the heteroscedasticity and autocorrelation consistent estimator (HAC; Newey and West, 1986), assuming a general dispersion pattern.

The multiple linear regression model is a well documented tool in statistics (e.g. Seber, 2003). If the usual LS assumptions are true, the generalized least squares estimator (GLSE; Greene, 2007) is the best linear unbiased estimator of the GMM with known VC-matrix within the class of linear estimators:

$$\hat{\mathbf{x}} = (\mathbf{A}'\mathbf{\Omega}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{\Omega}^{-1}\mathbf{l} \tag{9}$$

The GLSE will still be unbiased assuming that (i) the known VC-matrix $\mathbf{\Omega}_1$ is the first approximation of the non-spherical behavior with respect to error term and (ii) the true dispersion pattern of $\mathbf{l}$ is given by $\mathbf{\Omega}_0$, which should only belong to the set of all symmetric and positive definite matrices. This could be seen by using the expectation operator on $\hat{\mathbf{x}}$:

$$\begin{aligned}E(\hat{\mathbf{x}}) &= E\left((\mathbf{A}'\mathbf{\Omega}_1^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{\Omega}_1^{-1}\mathbf{l}\right) \\ &= (\mathbf{A}'\mathbf{\Omega}_1^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{\Omega}_1^{-1}\mathbf{A}\mathbf{x} \\ &= \mathbf{x}\end{aligned} \tag{10}$$

It is feasible to use the GLSE as long as the true parameter vector $\mathbf{x}$ has been estimated on the average, independent of any arbitrary approximation of the true VC-matrix of the population. Regarding to the alternative GLSE expression

$$\hat{\mathbf{x}} = \mathbf{x} + (\mathbf{A}'\mathbf{\Omega}_1^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{\Omega}_1^{-1}\mathbf{v} \tag{11}$$

the VC-matrix of $\hat{\mathbf{x}}$ can be calculated as follows

$$\begin{aligned}V(\hat{\mathbf{x}}) &= E\left[(\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})'\right] \\ &= E\left[(\mathbf{A}'\mathbf{\Omega}_1^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{\Omega}_1^{-1}\mathbf{v}\mathbf{v}'\mathbf{\Omega}_1^{-1}\mathbf{A}(\mathbf{A}'\mathbf{\Omega}_1^{-1}\mathbf{A})^{-1}\right] \\ &= (\mathbf{A}'\mathbf{\Omega}_1^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{\Omega}_1^{-1}\mathbf{\Omega}_0\mathbf{\Omega}_1^{-1}\mathbf{A}(\mathbf{A}'\mathbf{\Omega}_1^{-1}\mathbf{A})^{-1} \\ &\neq (\mathbf{A}'\mathbf{\Omega}_1^{-1}\mathbf{A})^{-1}\end{aligned} \tag{12}$$

As shown it Equation 12, the use of $\mathbf{\Omega}_1$ as the true VC-matrix of the population results in a wrong computation of the estimated parameters. The consequences of this failure within the esti-

mation process will lead to invalidated statistical joint or single hypotheses tests. The degree of complexity is further enhanced when both $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_0$ are unknown (this is the common situation in statistics) and have to be estimated from the data. The estimated GLSE (or two stage Aitken estimator) is given by:

$$\tilde{\mathbf{x}} = (\mathbf{A}'\hat{\mathbf{\Omega}}_1^{-1}\mathbf{A})^{-1}\mathbf{A}'\hat{\mathbf{\Omega}}_1^{-1}\mathbf{l} \tag{13}$$

This estimator is neither linear due the fact that $\hat{\mathbf{\Omega}}_1$ and $\mathbf{v}$ are correlated, nor does it have known finite sample properties in general. At least it can be said that $\tilde{\mathbf{x}}$ is a consistent estimator of $\mathbf{x}$ whenever the matrix series $(\hat{\mathbf{\Omega}}_1)_i^{\infty}$ converges in probability to $\hat{\mathbf{\Omega}}_0$ due Slutsky's theorem (Greene, 2007). For sure it is impossible to estimate $\mathbf{\Omega}_0$ directly because of the $0.5N(N+1)$ free parameters within that VC-matrix (N: Number of observations).

The authors adopt two well known robust covariance estimation procedures from econometrics to circumvent the direct estimation problem. The idea is quite simple: Seeking a consistent estimator for $\mathbf{A}'\mathbf{\Omega}_1^{-1}\mathbf{\Omega}_0\mathbf{\Omega}_1^{-1}\mathbf{A}$, instead of estimating a restricted version of the true VC-matrix of the population. This matrix contains only $0.5n(n+1)$ free parameters whereas n is referred the number of components in $\mathbf{x}$ (n: Number of unknown parameters). Assuming that $\mathbf{\Omega}_0$ is a positive definite diagonal matrix (without an autocorrelation pattern) and $\mathbf{\Omega}_1 = \mathbf{I}$ is the best initial approximation over the underlying HS pattern, the approach will coincide with White's heteroscedasticity consistent estimator (White, 1980). Furthermore, the only restrictions on $\mathbf{\Omega}_0$ are given by symmetry and positive definiteness, which lead to the heteroscedasticity and autocorrelation consistent estimator of Newey and West (1986). The estimation process is carried out in the following way:

1. Using Equation 13 whereas $\mathbf{\Omega}_1$ represents the stochastic model and the best approximation of $\mathbf{\Omega}_0$ ($\mathbf{\Omega}_1$: VC-matrix estimated by VCE).

2. Estimating the true VC-matrix of $\mathbf{x}$ by

$$V(\hat{\mathbf{x}})_{HC} = (\mathbf{A}'\hat{\mathbf{\Omega}}_1^{-1}\mathbf{A})^{-1}\tilde{\Sigma}_0(\mathbf{A}'\hat{\mathbf{\Omega}}_1^{-1}\mathbf{A})^{-1} \tag{14}$$

   with $\tilde{\Sigma}_0 = \mathbf{A}'\hat{\mathbf{\Omega}}_1^{-1}\tilde{\mathbf{\Psi}}_0\hat{\mathbf{\Omega}}_1^{-1}\mathbf{A}$
   $\tilde{\mathbf{\Psi}}_0 = Diag(\hat{v}_i^2)$

   and assuming a HC pattern as well as abstaining from any autocorrelation within the disturbances only.

3. Estimating the true VC-matrix of $\mathbf{x}$ by

$$V(\hat{\mathbf{x}})_{HAC} = (\mathbf{A}'\mathbf{\Omega}_1^{-1}\mathbf{A})^{-1}\tilde{\Sigma}_1(\mathbf{A}'\mathbf{\Omega}_1^{-1}\mathbf{A})^{-1} \tag{15}$$

   with $\tilde{\Sigma}_1 = \left[\tilde{\Sigma}_0 + \tilde{\Gamma}_1\right]$
   $\tilde{\Gamma}_1 = \sum_{j=1}^{p}\left(1 - \frac{j}{p+1}\right)$
   $\cdot \sum_{i=j+1}^{N}(\mathbf{a}_i v_i v_{i-j}\mathbf{a}'_{i-j} + \mathbf{a}_{i-j}v_{i-j}v_i\mathbf{a}'_i)$
   $p = floor\left(4(N/100)^{2/9}\right)$

   and assuming a general dispersion pattern. The row vector $\tilde{\mathbf{a}}_i$ is the i^{th} row of $\tilde{\mathbf{A}} := \mathbf{WA}$ where $\mathbf{W}$ contains the reciprocals of $\sqrt{\lambda_i}$ whereas $\sqrt{\lambda_i}$ is the i^{th} eigenvalue of $\hat{\mathbf{\Omega}}_1$. This is due to a spectral decomposition of $\hat{\mathbf{\Omega}}_1$ (extension of the HAC estimator on heterogeneous observations).

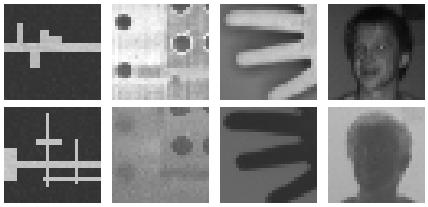
Since both VC-estimators are consistent for the true $\mathbf{\Omega}$ of the population (Newey and West, 1986; White, 1980) all single and joint hypotheses tests are valid for the asymptotic case.

5 RESULTS

The following section presents results of several experiments and shows the effects of different functional and stochastic LST models.

5.1 Experiment 1: Functional Model

To show the improved parameter accuracies and reliability of the estimation, simulated and real data (Figure 3) with (i) high intensity contrast, (ii) high range contrast and (iii) balanced contrast between both channels have been used. Furthermore, trials with and without a range offset were performed. This configuration should show the influence of different functional models on the shift and scale parameters.



(a) (b) (c) (d)

Figure 3: Some 3-D camera intensity (top) and range (down) images. (a): Synthetic data. (b): Real data with measuring marks. (c): Human hand. (d): Human head.

A synthetic data set has been created in order to determine the influences of different functional models accurately. These data are simply noisy ($\sigma_{gv} = 1000\,\text{bit}$; $\sigma_{rv} = 100\,\text{bit}$), but no systematic errors occur (e.g. measurement uncertainties due to surface conditions, background illumination or multipath effects). Real data have been captured by the *SwissRanger SR-3000* with constant integration time ($it = 20.2\,\text{ms}$) and modulation frequency ($mf = 20.0\,\text{MHz}$). After exploring the potential of 2.5-D LST with synthetic data, those experiments reflect the results, which can be expected in practical use.

For single channel estimation it is obvious that insufficient contrast in intensity or range observations will result in non-convergence of the solution vector. This is a general problem of LSM: As the covariance matrix is generated from observations with stochastic properties, the estimated standard deviations (SD) of the transformation shift components will generally be too optimistic, and correlations between parameters, indicating singularities caused by insufficient patch-gradients will often not be detected (Maas, 2002).

Consequently the adjustment is stabilized, if one of the channels provides sufficient contrast in both coordinate directions. Even though it contains low information, the corresponding intensity or range signal is not discarded but has some influence on the adjustment in the form of a slight improvement of shift and scale parameters against single channel estimation with high contrast (approx. 5.0 % for $\sigma_{a0,b0}$ and 20.0 % for $\sigma_{a1,b2}$). Therewith, the entire available information is used.

An improvement in scale adjustment can be achieved for significant depth variations between template and search patch, independent of even or uneven range patches (approx. 50.0 % for $\sigma_{a1,b2}$).

An optimal 2.5-D LST solution can be achieved, if intensity and range channel provide sufficient contrast. The SD for shift parameters $\sigma_{a0,b0}$ are within a range of 1/50 to 1/25 pixel for real data. A range offset can be determined with a relative accuracy of 0.25 % of the whole distance d_0.

A measure of the quality of the introduced functional model provide the SD of the adjusted observations $\hat{\sigma}_{gv,rv}$. Those values range around 100 bit for the intensity channel and 10 bit for the range one. Specific details on the order of magnitude (16 bit-range) will be given in Section 5.2.

σ_{a0} [pixel]	σ_{b0}	σ_{a1}	σ_{b2}	σ_{d0} [mm]	σ^0_{gv} [16bit]	σ^0_{rv}	σ_{gv} [16bit]	σ_{rv}	$\hat{\sigma}_{gv}$ [16bit]	$\hat{\sigma}_{rv}$	Comments
0.037	0.026	0.008	0.007	4.160	1.0	1.0	×	×	65.112	37.669	Equal weights
0.036	0.026	0.010	0.007	5.742	1.0	1·E+7	×	×	69.930	51.810	Overemphasis intensity measurement
0.069	0.054	0.007	0.007	0.475	1·E+7	1.0	×	×	103.232	7.616	Overemphasis range measurement
0.041	0.034	0.006	0.006	0.641	1.0	1.0	1616.916	120.719	69.108	7.499	VCE (two groups of observation)

Table 1: RIM sequence analysis using different stochastic models. Experimental trials performed with 2.5-D LST using intensity and range channel.

$tr(V(\hat{\mathbf{x}})_{VCE})$	$tr(V(\hat{\mathbf{x}})_{HC})$	$tr(V(\hat{\mathbf{x}})_{HAC})$	$\frac{tr(\ldots)^*}{tr(\ldots)^{**}}$	$\frac{tr(\ldots)^*}{tr(\ldots)^{***}}$	R	Homogeneity *,**	Homogeneity *,***	Comments
0.00349^*	0.00492^{**}	0.00550^{***}	0.709	0.634	[0.10 ; 9.60]	☑	☑	Static scene Measuring mark Images taken form different point of views
0.00688^*	0.01696^{***}	0.01234^{**}	0.557	0.405	[0.10 ; 9.60]	☑	☑	
0.00260^{***}	0.00257^{**}	0.00256^*	0.986	0.999	[0.14 ; 7.15]	☑	☑	Kinematic scene Natural gray value gradient Images taken form one point of view over time
0.05099^{**}	0.04448^*	0.12381^{***}	0.872	0.359	[0.14 ; 7.15]	☑	☑	

Table 2: Homogeneity of variance covariance matrices.

The experiment shows that the use of complementary information improve accuracy and reliability for RIM matching tasks. 2.5-D LST is most impressive when dealing with significant range offsets between template and search patches. In particular the range channel supports gray value observations in scale adjustment and provides additional information in the case of low contrast within intensity patches.

5.2 Experiment 2: Stochastic Model

When processing heterogeneous data, an adaption of the stochastic model is necessary. A splitting of a single heterogeneous observation group in several ones allows the consideration of multiple variance factors (Section 3.2). Thus, it is possible to tap the full information potential of the available observations. The results of some experiments on a RIM data set with high intensity and range contrast as well as a range offset of about 30 cm between template and search patch accentuate the need for an adapted stochastic model (Table 1). The following experimental setup was used: 2.5-D LST with (i) equal weights for intensity and range observations ($\sigma^0_{gv,rv} = 1\,\text{bit}$), an overemphasis of (ii) the intensity ($\sigma^0_{rv} = 1 \cdot \text{E} + 7\,\text{bit}$) or (iii) the distance measurement ($\sigma^0_{gv} = 1 \cdot \text{E} + 7\,\text{bit}$) and (iv) a stochastic model, which was estimated by VCE with $\sigma^0_{rg} = 1\,\text{bit}$ and $\sigma^0_{rg} = 1\,\text{bit}$ as initial a-priori SD for a VCE.

In Table 1 (Row 1-3) a negative influence on parameter accuracies is obvious for a deficient weighting of the measurements. The specified a-posteriori SD of the adjusted observations have higher variances, compared to a well-balanced weighting (Row 4). Those balanced weighting could be achieved by VCE with two groups of observations. The precision of the shift parameters is within the order of 1/30 pixel. The scales can be determined with a precision $\sigma_{a1,b2}$ of 0.006, and the corresponding range offset SD σ_{d0} (derived from Equation 7 by the law of the propagation of errors) becomes 0.6 mm (based on $d_0 = 294.1\,\text{mm}$, corr. 0.2 %).

Applying a VCE, precision information of the original observations becomes available: In this case, the a-priori SD of the intensity measurement σ_{gv} is 1600 gray value, which corresponds to 16 bit resp. 6 gray values referring to 8 bit. This magnitude is realistic and comprehensible due to a poorer signal-to-noise ratio of CMOS sensors (Lange, 2000), background illumination and variations within charge-to-voltage relation. Furthermore, σ_{gv} aligns with previous results empirically determined by Hempel (2006). The range values have been measured with an SD of 121 counts resp. 1.4 cm ($it = 20.2\,\text{ms}; mf = 20.0\,\text{MHz}$), which corresponds to Hempel's results as well. Free of systematic errors, the averaged a-posteriori SD of the adjusted observations $\hat{\sigma}_{gv,rv}$ can be specified with 69 counts (16 bit-range) resp. 0.3 gray values (8 bit-range) for the intensity channel and 8 counts resp. 1 mm for the range channel. Those values reflect that the raw data match well with the established model.

The experiment shows that 2.5-D LST in combination with a VCE improves the parameter accuracies. Especially if no adequate a-priori precision information is available, an optimal utilization of the whole content of information becomes possible. Furthermore, the procedure delivers valuable information on the sensor and data quality.

5.3 Experiment 3: Robust VC-Matrix Estimation

The following experiments have been performed to show, whether an enhanced stochastic model in the form of a robust VC-matrix estimation (Section 4) is useful for the presented functional model (Section 3.2) and to quantify differences in the precision of the underlying VC-estimators (VCE, HC and HAC). Testing the determinants of the estimated VC-matrices of the GMM parameter vector $\hat{\mathbf{x}}$ would be of a great benefit since this approach will incorporate the covariances between the components of $\hat{\mathbf{x}}$ as well as their variances. However, this test procedure will only work for orthogonal regressions or stochastic processes (Rounault, 2007), which both do not fit into the presented framework. In order to still derive a valid value for the homogeneity of the several VC-matrices, the traces $tr(V(\hat{\mathbf{x}})_{VCE})$, $tr(V(\hat{\mathbf{x}})_{HC})$ and $tr(V(\hat{\mathbf{x}})_{HAC})$ of the estimated matrices are tested against each other by usual F-test procedures:

- Null hypothesis:
$$H_0 : \sigma_1^2 = \sigma_2^2 \tag{16}$$

- Test Statistics:
$$T = \frac{tr\,(V(\hat{\mathbf{x}})_*)}{tr\,(V(\hat{\mathbf{x}})_{**})} \quad \text{resp.} \quad \frac{tr(V(\hat{\mathbf{x}})_*)}{tr(V(\hat{\mathbf{x}})_{***})} \tag{17}$$
with $tr\,(V(\hat{\mathbf{x}})_*) < tr\,(V(\hat{\mathbf{x}})_{**}) < tr\,(V(\hat{\mathbf{x}})_{***})$

- Acceptance region:
$$R = \left[F_{n-1,n-1,\frac{\alpha}{2}}; F_{n-1,n-1,1-\frac{\alpha}{2}} \right] \tag{18}$$
with F: Quantile of F-distribution
n: Number of unknown parameters
α: Significance level

Table 2 shows the results of variance homogeneity tests for some representative 2.5-D LST tasks. Obviously, the VC-matrices are homogeneous; this means that no significant variations occur and the VC-matrix estimated by VCE can be accepted. In other words, when using the same sensor, the measurements of one group of observations (intensity or range) is subject to the same stochastic errors. Thus, an aggregation of all observations of one group with one weight is acceptable. An a-posteriori weighting by robust VC-matrix estimation is not necessary. Due to its low computational effort, the estimation of robust VC-matrices is still practicable and maybe useful in the case of VCE modeling failures within the VC-matrix.

The experiment shows that the consideration of two groups of observations in a VCE is sufficient and that there is no significant variation of the precision of observations within the groups of observations.

6 CONCLUSIONS AND OUTLOOK

In this article, a least squares tracking approach based on 3-D camera intensity and range data was proposed. The presented functional model combines the transformation parameters for intensity and range images and has been proven by various experiments with synthetic and real data. It could be shown that an increase in accuracy, stability and reliability can be reached for least squares matching and tracking by the integrated treatment of intensity and range information. The stochastic model has been designed by using a variance component estimation approach as well as robust variance covariance matrix estimation. It could be shown that a separation of the heterogeneous data into two groups of observations is sufficient for the accuracy of the stochastic model, and that there is no significant variation of precision within the groups of observations. As an additional product, the procedure delivers information on the precision of 3-D camera range and intensity measurements.

So far, the affine scale parameters are modeled through the additional range information. Future work will address other geometric patch transformation parameters, which are not considered by the 2-D affine transformation. In particular, the keystone distortion caused by an inclination between the sensor plane and the captured object can be expressed through the RIM depth offset parameter. Beyond this, the effect and elimination of outliers should be addressed. Robust estimation procedures with respect to outlier phenomena will be more appropriate than dealing with certain HS-patterns. Also, likelihood approaches with a heavy-tailed distribution could be used for a Gaussian distributed error term within the GLSE framework.

References

Ackermann, F., 1984. High precision digital image correlation. In: Proceedings of the 39th Photogrammetric Week, Schriftenreihe der Universität Stuttgart, Vol. 9, pp. 231–243.

Baltsavias, E. P., 1991. Multiphoto Geometrically Constrained Matching. Ph.d. thesis, Institute of Geodesy and Photogrammetry. ETH Zurich. Switzerland.

Breuer, P., Eckes, C. and Muller, S., 2007. Hand gesture recognition with a novel ir time-of-flight range camera: A pilot study. In: MIRAGE07, pp. 247–260.

Du, H., Oggier, T., Lustenberger, F. and Charbon, E., 2005. A Virtual Keyboard Based on True-3D Optical Ranging. In: British Machine Vision Conference 2005, Vol. 1, pp. 220–229.

Förstner, W., 1984. Quality assessment of object location and point transfer using digital image correlation techniques. In: International Archives of Photogrammetry, Vol. 25-III, A3a, pp. 197–217.

Göktürk, S. B. and Tomasi, C., 2004. 3d head tracking based on recognition and interpolation using a time-of-flight depth sensor. In: CVPR04, pp. II: 211–217.

Greene, W. H., 2007. Econometric Analysis. 6th edn, Prentice Hall. ISBN 978-0130661890.

Grün, A., 1985. Adaptive least squares correlation - a powerful image matching technique. In: South African Journal of Photogrammetry, Remote Sensing and Cartography, Vol. 14, 3, pp. 175–187.

Grün, A. and Akca, D., 2004. Least squares 3d surface matching. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 5, Vol. XXXIV, ISPRS "Panoramic Photogrammetry Workshop".

Gudmundsson, S. A., 2006. Robot vision applications using the CSEM swissranger camera. Master's thesis, Informatics and Mathematical Modelling, Technical University of Denmark, DTU.

Hempel, M., 2006. Validierung der Genauigkeit und des Einsatzpotentials einer distanzmessenden Kamera. Unveröffentlichte Diplomarbeit, Technische Universität Dresden Institut für Photogrammetrie und Fernerkundung.

Hempel, R. and Westfeld, P., 2008. Statistical modeling of interpersonal distance with range imaging data. In: Accepted at COST 2102 training school on "Multimodal Signals: Cognitive and Algorithmic Issuses". Vietri sul Mare, Italy, 21st-26th April 2008.

Holte, M. B. and Moeslund, T. B., 2007. Gesture recognition using a range camera. Technical Report CVMT-07-01 ISSN 1601-3646, Aalborg University, Denmark.

Isard, M. and Blake, A., 1998. Condensation – conditional density propagation for visual tracking. International Journal of Computer Vision 29(1), pp. 5–28.

Kahlmann, T. and Ingensand, H., 2006. Calibration of the fast range imaging camera swissranger for use in the surveillance of the environment. In: Electro-Optical Remote Sensing II. Edited by Kamerman, Gary W.; Willetts, David V.; Steinvall, Ove K.. Proceedings of the SPIE, Volume 6396, pp. 639605 (2006)., Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, Vol. 6396.

Kahlmann, T., Remondino, F. and Guillaume, S., 2007. Range imaging technology: new developments and applications for people identification and tracking. In: A. Beraldin, F. Remondino and M. R. Shortis (eds), Proceedings of Videometrics IX - SPIE-IS&T Electronic Imaging, Vol. 6491.

Kahlmann, T., Remondino, F. and Ingensand, H., 2006. Calibration for increased accuracy of the range imaging camera swissrangertm. In: H.-G. Maas and D. Schneider (eds), ISPRS Image Engineering and Vision Metrology, Vol. XXXVI, 5, pp. 136–141. ISSN 1682-1750.

Koch, K.-R., 2004. Parameterschätzung und Hypothesentests in linearen Modellen. 4. edn, Ferd. Dümmlers Verlag, Bonn.

Koller-Meier, E. B., 2000. Extending the condensation algorithm for tracking multiple objects in range image sequences. PhD thesis, Technische Wissenschaften ETH Zrich, Nr. 13548.

Lange, R., 2000. 3D Time-of-Flight Distance Measurement with Custom Solid-State Image Sensors in CMOS/CCD-Technology. Phd thesis, Department of Electrical Engineering and Computer Science at University of Siegen.

Maas, H.-G., 2000. Least-squares matching with airborne laserscanning data in a tin structure. International Archives of Photogrammetry and Remote Sensing 33(3a), pp. 548–555.

Maas, H.-G., 2002. Methods for measuring height and planimetry discrepancies in airborne laserscanner data. Photogrammetric Engineering & Remote Sensing 68(9), pp. 933–940.

Maas, H.-G., Stefanidis, A. and Grün, A., 1994. From pixels to voxels - tracking volume elements in sequences of 3-d digital images. In: International Archives of Photogrammetry and Remote Sensing, 3/2, Vol. 30.

Newey, W. K. and West, K. D., 1986. A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix. NBER Technical Working Papers 0055, National Bureau of Economic Research, Inc.

Oggier, T., Lehmann, M., Kaufmann, R., Schweizer, M., Richter, M., Metzler, P., Lang, G., Lustenberger, F. and Blanc, N., 2004. An all-solid-state optical range camera for 3D real-time imaging with sub-centimeter depth resolution (SwissRanger). In: L. Mazuray, P. J. Rogers and R. Wartmann (eds), Optical Design and Engineering, Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, Vol. 5249, pp. 534–545.

Rounault, A., 2007. Asymptotic behavior of random determinants in laguerre, gram and jacobi ensembles. Latin American Journal of Probability and Mathematical Statistics III, pp. 181–230.

Schneider, D. and Maas, H., 2007. Integrated bundle adjustment with variance component estimation: Fusion of terrestrial laser scanner data, panoramic and central perspective image data. In: Proceedings ISPRS Workshop Laser Scanning and SilviLaser 2007, Helsinki, International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVInumber 3/W52, p. 373.

Schwarte, R., 1997. Verfahren und Vorrichtung zur Bestimmung der Phasen- und/oder Amplitudeninformation einer elektromagnetischen Welle. Patentschrift DE 19704496, Inhaber Schwarte GBR.

Seber, G. A. F., 2003. Linear Regression Analysis (Wiley Series in Probability and Statistics). 1st edn, John Wiley & Sons, Inc. ISBN 978-0471415404.

Westfeld, P., 2007a. Ansätze zur Kalibrierung des Range-Imaging-Sensors SR-3000 unter simultaner Verwendung von Intensitäts- und Entferungsbildern. In: T. Luhmann (ed.), Photogrammetrie - Laserscanning - Optische 3D-Messtechnik (Beiträge Oldenburger 3D-Tage 2007), Herbert Wichmann Verlag Heidelberg, pp. 137–146.

Westfeld, P., 2007b. Development of approaches for 3-d human motion behaviour analysis based on range imaging data. In: A. Gruen and H. Kahmen (eds), Optical 3-D Measurement Techniques VIII, Vol. II, Institute of Geodesy and Photogrammetry, ETH Zürich, pp. 393–402.

White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48(4), pp. 817–838.

Zywitza, F., Massen, J., Brunn, M., Lang, C. and Görnig, T., 2005. One-to-three-dimensional ranging for future automotive safety systems. In: H. Ingensand and T. Kahlmann (eds), Proceedings of the 1st Range Imaging Research Day at ETH Zurich in Switzerland. ISBN 3-906467-57-0.