

# Has Something Changed Here?

## Autonomous Difference Detection for Security Patrol Robots

Henrik Andreasson, Martin Magnusson and Achim Lilienthal

Centre of Applied Autonomous Sensor Systems, Dept. of Technology, Örebro University, Sweden

henrik.andreasson@tech.oru.se, martin.magnusson@tech.oru.se, achim@lilienthals.de

**Abstract**—This paper presents a system for autonomous change detection with a security patrol robot. In an initial step a reference model of the environment is created and changes are then detected with respect to the reference model as differences in coloured 3D point clouds, which are obtained from a 3D laser range scanner and a CCD camera. The suggested approach introduces several novel aspects, including a registration method that utilizes local visual features to determine point correspondences (thus essentially working without an initial pose estimate) and the 3D-NDT representation with adaptive cell size to efficiently represent both the spatial and colour aspects of the reference model. Apart from a detailed description of the individual parts of the difference detection system, a qualitative experimental evaluation in an indoor lab environment is presented, which demonstrates that the suggested system is able register and detect changes in spatial 3D data and also to detect changes that occur in colour space and are not observable using range values only.

### I. INTRODUCTION

An important aspect of robotic security systems is surveillance of a specified area. Typical security patrol missions require detection of changes in the environment and description of detected differences with respect to a previously determined reference state. A human watchman is first shown around the premises to learn the reference state of the environment. The mission of the security patroller is then to check for changes in the environment, e.g. looking for open doors, open windows, water leaks, blocking of fire hoses etc. [1]. Due to their dull and potentially dangerous character it is desirable to delegate security patrols to mobile robots. Current approaches often require teleoperation [2]. Accordingly, research has focused on adjustable autonomy or semi-autonomy to decrease the amount of “cognitive burden” to the operator (Seeman et. al [3], Goodrich et. al [4], for example).

This paper presents a system for autonomous change detection with a security patrol robot. As for its human counterpart the robot watchman is expected to determine the reference state of the environment (the reference model) in an initial phase. The actual mission requires discovering changes with respect to the reference model and describing the differences. An alarm may be triggered whenever changes are detected or detected differences can be reported to a human operator for further analysis. An effective system for autonomous difference detection needs to fulfil a couple of demands. First, creation of the reference model should be simple and require minimal effort by a user. Ideally the

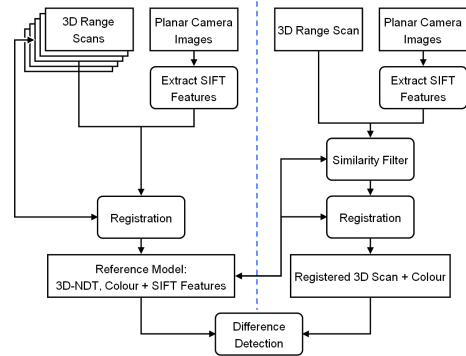


Fig. 1. Overview of the difference detection system. Acquisition of the reference model is shown on the left side of the dashed line and pre-processing of new data for the actual difference detection to the right of the dashed line.

robot would acquire and update the reference model of the environment autonomously. This is related to the dynamic mapping problem that investigates continuous adaptation of maps over time [5]. Here we assume that the mobile security robot is first guided by a human operator in order to learn a model of the original, unmodified environment. A second major requirement especially with regards to large environments is that the reference model is represented efficiently but nevertheless allows determining small changes in the environment. Representing the environment at a high resolution so as to avoid the need to inspect all parts of the environment from a small distance can be seen generally as the third major requirement of an autonomous difference detection system.

Difference detection has been studied in the context of recognizing parked vehicles by Ishikawa et. al [6]. In their work, Ishikawa et. al use an omni-directional camera together with GPS and INS (Inertia Navigation System) to obtain depth estimates, which were subsequently compared with a pre-built model.

### II. METHOD

An overview of the suggested difference detection system is shown in Fig. 1. Corresponding to the two columns in the figure, the approach decomposes into two parts: acquisition of the reference model (shown to the left of the dashed line) and pre-processing of new data (shown to the right of the dashed line) for the actual detection of differences between

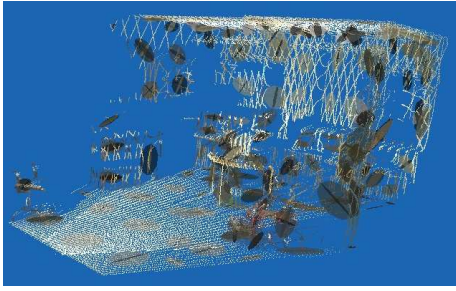


Fig. 3. Visualization of the 3D-NDT representation corresponding to the scan in Fig. 2, top. Ellipsoids indicate the eigenvalues of the respective covariance matrices.

new data and the reference model (indicated by the box below the dashed line).

Our difference detection approach comprises three main components described in Sec. III – Sec. V below. A very important aspect is accurate registration of the data. We present a registration method that produces accurate results even under the condition that a reasonable relative pose estimate is not available (when there is essentially no estimate of the relative pose between two data sets other than they were recorded in the same environment). This is achieved by using local visual features extracted from the camera images to solve the data association problem in the registration process (Section III-A – III-B). Since the camera images provide a much higher resolution than the range sensor our method requires that depth information is estimated at the position of local visual features in the image (Section III-C). To calculate depth information from a low-resolution 3D range sensor and a colour image, we use a non-iterative method for vision-based interpolation of 3D range scans that was introduced by the authors in an earlier paper [7]. Finally a modified closed-form solution is used for the actual registration (Section III-D).

In order to represent the environment efficiently, we apply the Normal Distribution Transformation (NDT) with adaptive cell splitting (Section IV-A) to the spatial point distribution and the colour distribution (Section IV-C and Section IV-C). Finally, based on the 3D-NDT, respectively the Colour 3D-NDT representation, difference probabilities are calculated as described in Sec. V-A and V-B.

A visualisation of the difference probabilities in a fairly trivial case where two 3D laser scan sweeps are compared can be seen Fig. 2. The corresponding 3D-NDT representation is shown in Fig. 3 using ellipsoids to indicate the eigenvalues of the respective covariance matrices.

### III. REGISTRATION

A very important aspect of the difference detection system is that the relative position  $[\mathbb{R}, \mathbf{t}]$  between the current  $\mathcal{S}_c$  and previous scans  $\mathcal{S}_p$  is known. Accurate registration is therefore a fundamental requirement. Since the measurements (3D range scans and colour images) to be compared will be obtained not only at different poses but also at substantially different times, registration should be robust to a certain

level of changes in the environment. In addition, it cannot be expected in general that the security patrol robot maintains a consistent coordinate system in-between acquisition of the reference model and a difference detection request. As opposed to most current approaches to scan registration we therefore need a registration method that does not depend on reasonably accurate initial pose estimates.

In order to cope with the condition that initial pose estimates may not be available, we use local visual features to establish correspondences between data points (data association) within the entire set of measurements. In our current implementation we use the popular SIFT features developed by Lowe [8].

Shortly the registration procedure can be described as follows: first, SIFT features are computed in the planar images recorded with the current scan data  $\mathcal{S}_c$  and compared to the SIFT features found in the images belonging to previous scan. As detailed in Section III-A, a similarity measure is calculated from the number of matching features, which allows to determine the scan in the reference model  $\mathcal{S}_r^*$  that is most similar to the current scan  $\mathcal{S}_c$  (see Section III-B). Next, the depth values are estimated for all matching feature pairs in  $\mathcal{S}_r^*$  and  $\mathcal{S}_c$ , using vision-based interpolation as described in Section III-C. The pairs of 3D points corresponding to matching features are then used in a closed form solution to obtain the relative pose estimate (see Section III-D).

In a related approach Newman et. al. [9] used SIFT features to detect loop closure events in a 3D SLAM approach. In contrast to their method where SIFT features are used to obtain an initial pose estimate (by determining the essential matrix between two images) and the full point cloud is considered afterwards, registration in our approach is carried out using only 3D points that are associated with matching visual features. By restricting scan matching to 3D points that were found to correspond by their visual appearance, we believe that the robustness against changes in the environment is improved and more accurate registration can be obtained. However, it remains to validate this belief in a thorough ground truth evaluation using a large set of 3D scans.

#### A. Appearance-Based Similarity Measure

Given two images  $I_a$  and  $I_b$ , we extract local visual features using the SIFT algorithm [8] resulting in two sets of features  $F_a$  and  $F_b$ , corresponding to the two images. Each feature  $f_i = \{[X, Y]_i, H_i\}$  in a feature set  $F = \{f_i\}$  comprises the position  $[X, Y]_i$  in pixel coordinates and a histogram  $H_i$  containing the SIFT descriptor. The similarity  $S_{a,b}$  of image  $I_a$  and  $I_b$  is based on the number of matching features between  $F_a$  and  $F_b$ .

The feature matching algorithm calculates the Euclidean distance between each feature in image  $I_a$  and all the features in image  $I_b$ . A potential match is found if the smallest distance is smaller than 60% of the second smallest distance. This criterion was found empirically and was also used in [10], for example. It reduces the risk of falsely declaring

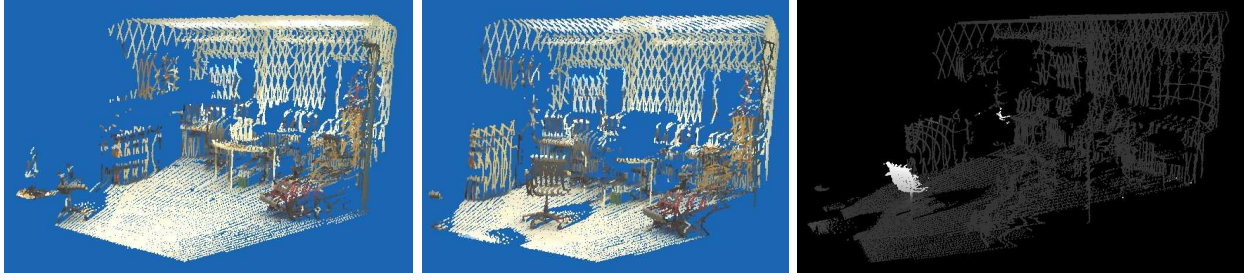


Fig. 2. Left, middle: 3D range scans recorded at a relatively small displacement of approx. 1 m. Right: Difference probability corresponding to a chair that was placed in the scene before the second scan was recorded. Brighter regions indicate a higher difference probability.

correspondence between SIFT features by excluding cases where a false correspondence is caused by the existence of several almost equally well matching alternatives. In addition, no feature is allowed to be matched against more than one other feature. If more than one candidate for matching is found, the feature with the highest similarity among the candidate matches is selected.

The feature matching step results in a set of feature pairs  $P_{a,b}$ , with a total number  $M_{a,b} = |P_{a,b}|$  of matched pairs. Since the number of extracted features varies heavily depending on the image, the number of matches is normalized to the average number of features in the two images. Hence the similarity measure  $S_{a,b} \in [0, 1]$  is defined as:

$$S_{a,b} = \frac{M_{a,b}}{\frac{1}{2}(n_{F_a} + n_{F_b})} \quad (1)$$

where  $n_{F_a} = |F_a|$  and  $n_{F_b} = |F_b|$  are the number of features in  $F_a$  and  $F_b$  respectively.

### B. Similarity Filter

High similarity values correspond to a similar appearance of the respective images and can be taken as an indication of a similar position at which the images were recorded. In the difference detection system suggested in this paper we apply pair wise scan-matching. After the reference model was built and a current scan was recorded at a certain pose it therefore needs to be decided which scan in the reference model the current scan should be registered with. This decision is made based on the similarity measure in Eq. 1. For the current scan  $\mathcal{S}_c$ , the most similar scan in the reference model  $\mathcal{S}_r^*$  is selected for registration. Since each 3D scan is associated with three range scanner sweeps and seven planar images the scan in the reference model that is most suitable for registration is selected as the one that maximises the total similarity. The total similarity of two scans is calculated by matching all images belonging to one scan to all images belonging to the other scan and considering the sum of features and feature matches in Equation 1. This approach corresponds to matching two combined images which were created by stitching the individual images recorded with each scan. Note that the similarity filter is not required in the initial model acquisition where an initial estimate of the relative pose between subsequent scans is provided by the odometry

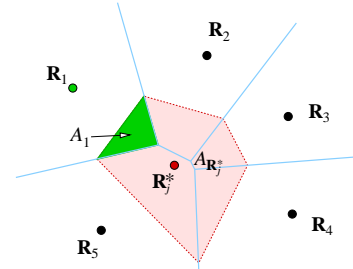


Fig. 4. Natural neighbours  $\mathbf{R}_1..R_5$  of  $\mathbf{R}_j^*$ . The interpolated weight of each natural neighbour  $\mathbf{R}_i$  is proportional to the size of the area which contains the points' Voronoi cell and the cell generated by  $\mathbf{R}_j^*$ . I.e. the nearest neighbour  $\mathbf{R}_1$  will have influence based upon the area of  $A_1$ .

of the robot. Here each 3D scan is registered to the previous scan.

### C. Depth Interpolation Using Vision

To obtain the depth estimate  $r_j^*$  for SIFT feature  $f_j$  the PLIC (Parameter-Free Multi-Linear Interpolation Considering Colour) method [7] is applied, which is briefly described below. Based on the observation that depth discontinuities in a scene often correspond to colour or brightness changes in the camera image, the PLIC method derives high-resolution depth images from a low-resolution 3D range scan and a colour image. In the difference detection system presented here, the PLIC method is used to determine depth values for matched SIFT features.

Image data consist of a set of image pixels  $\mathcal{P}_j = (X_j, Y_j, C_j)$ , where  $X_j, Y_j$  are the pixel coordinates and  $C_j = (C_j^1, C_j^2, C_j^3)$  is a three-channel colour value. By projecting a 3D laser reading point  $p_i = [x, y, z]$  with the range  $r_i$ , onto the image plane, a projected laser range reading point  $\mathbf{R}_i = (X_i, Y_i, r_i, (C_i^1, C_i^2, C_i^3))$  is obtained, which associates a range value  $r_i$  with the coordinates and the colour of an image pixel. The interpolation problem can now be stated for a given pixel  $\mathcal{P}_j$  and a set of projected laser range readings  $\mathbf{R} = \mathbf{R}_i$ , as to estimate the interpolated range reading  $r_j^*$  as accurately as possible. We denote an interpolated point by  $\mathbf{R}_j^* = (X_j, Y_j, r_j^*, C_j^1, C_j^2, C_j^3)$ , see Fig. 4.

Depth estimates  $r_j^*$  are computed by weighted interpolation between the natural neighbours to the pixel position of feature  $f_i$ . The natural neighbours are a set of projected 3D laser reading points  $\mathbf{R}_i$  and their weights are determined by

an area of influence  $A_i$  calculated from the Voronoi diagram of the projected laser scan readings as shown in Fig. 4. In addition, the weights for the natural neighbours depend on the colour variance of the assigned area  $A_i$  and on the difference of the colour  $C_j$  of the interpolated point to the colour  $C_i$  of neighbouring points.

#### D. Trimmed Closed Form Solution

Since visual features are used to establish corresponding scan points, no further means of data association, such as searching for closest data points in ICP [11], [12], is necessary. Hence it is possible to obtain the relative pose estimate using a closed form solution. Although the SIFT features were found to be very discriminative (see for example [13]), there is of course still a risk that some of the correspondences are not correct. To further decrease the possibility of erroneous point associations, only a set fraction of the correspondences with the smallest spatial distance between corresponding points is used for registration. In the experiments presented in this paper the fraction was set to 70%. To estimate the relative rotation  $\mathbb{R}$  and relative translation  $\mathbf{t}$  the singular value decomposition (SVD) approach suggested by [14] is applied. Because the fraction of data points that is used to estimate the relative pose  $[\mathbb{R}, \mathbf{t}]_i$  between two scans depends on the previous estimate  $[\mathbb{R}, \mathbf{t}]_{i-1}$  (since the relative pose estimate affects the spatial distance between corresponding points), the SVD method needs to be applied in an iterative manner. Thus relative pose updates are calculated repeatedly with the SVD method using the previous estimate  $[\mathbb{R}, \mathbf{t}]_{i-1}$  as input to the next iteration step until a stopping criterion is met. Any initial pose estimate can be used (in the experiment presented in this work, we start with the assumption of an identical pose). As the stopping criterion in the experiments in this paper we used if the change of the mean squared error (MSE) of the spatial distance between the corresponding points compared to the previous iteration were less than  $10^{-6} m^2$ .

#### IV. NORMAL DISTRIBUTION TRANSFORM (3D-NDT)

One of the major requirements for an autonomous difference detection system is that it scales well with the size of the environment. This demands an efficient representation of the reference model that compresses the data (to be able to store and maintain large environments) and is yet able to represent small details allowing for detection of small changes. To address this issue, the normal distribution transform (NDT) is used to represent the environment. The NDT was introduced by Biber et. al [15] and first used to register 2D laser scans. The method has been extended to 3D scan registration by Magnusson et. al. [16] and very recently, a multi-resolution variant was presented by Magnusson et. al. [17]. However, in this work we use a different approach to registration, as discussed in Section III.

The basic principle of NDT and 3D-NDT is to represent the environment by using a set of Gaussian distributions. First, the considered space is divided into cells. Each cell

that contains a certain minimum number of points is represented by a Gaussian with the mean value and covariance matrix computed from the points in the respective cell. The minimum number was set to 5 in the experiments presented in this paper. More formally 3D-NDT can be described as follows. We consider a point cloud  $\mathbf{P} = p_1, p_2, \dots, p_n$  with points  $p = [x, y, z]$  given in 3D Cartesian coordinates. The environment is divided into a set of cells  $\mathbf{C} = c_1, c_2, \dots, c_k$  and for each cell  $c_i$  a reference is stored to all  $N_c$  points  $p_{c_i}$  which lie within the cells boundaries. Then, for each cell  $c$ , the mean  $\mu_c$  and covariance  $C_c$  are calculated as

$$\mu_c = \frac{1}{N_c} \sum_{i=1}^{N_c} p_{c_i} \quad (2)$$

$$C_c = \frac{1}{N_c - 1} \sum_{i=1}^{N_c} (p_{c_i} - \mu_c)^2 \quad (3)$$

Note that the NDT representation does not require evenly spaced data and hence can be calculated without further sub-sampling. In order to keep the storage requirements within the limits of the computer with which the computations were made (i.e. 512 MB of RAM memory), all scans were sub-sampled with a 3D-grid resolution of  $0.1 m^3$ .

#### A. Adaptive Cell Splitting

The level of detail that is maintained by the NDT depends on the chosen cell size. In order to select the resolution of the NDT representation according to the local content of a scene, we use an adaptive cell size approach that determines whether or not to split a cell according to its covariance matrix  $C$ . From the picture of the covariance matrix as an ellipsoid, it is clear that a single Gaussian, i.e. one covariance matrix, can efficiently describe planes (one of the ellipsoid axis is small) and lines (two of the ellipsoid axis are small). Therefore, a large volume  $v_c$  of this ellipsoid, indicating that none of the principal axis is small, was chosen as the criterion for cell splitting:

$$v_c = |\lambda_1| |\lambda_2| |\lambda_3|, \quad (4)$$

where  $\lambda_i$  are the different eigenvalues of the covariance matrix. A higher value indicates a higher need to divide the cell. In the experiments presented in this work, the splitting threshold was set to  $0.001 m^3$ .

The decision about where to split a cell is made using the mean value  $\mu_c$  and the direction of the eigenvector  $e_c^{max}$  with the highest eigenvalue  $\lambda$ . The points  $p_{c_i}$  in the cell  $c_i$  that is to be split are assigned to new cells  $c_i^a$  and  $c_i^b$  according to the sign of

$$p_{c_i} (e_c^{max} - \mu_c). \quad (5)$$

Consequently, the cell is split at the plane through the centre  $\mu_c$  and orthogonal to the largest eigenvector  $e_c^{max}$ .

## B. Colour 3D-NDT

In order to be able to detect changes that do not reveal themselves through sufficiently large changes in the range readings but which are observable in the planar images recorded along with the range readings, colour information is also incorporated into the NDT representation. This allows detecting changes caused by thin objects, for example a poster that has been removed from the wall, as long as the colour differs sufficiently from the background.

In addition to points  $p = [x, y, z]$  in 3D space, we now also consider corresponding colour values  $\check{p} = [C^1, C^2, C^3]$ . Accordingly, the cells in the colour 3D-NDT representation are described by a mean value  $\check{\mu}$  and covariance  $\check{C}$  in addition to the spatial mean  $\mu$  and covariance  $C$  introduced above. The colour mean  $\check{\mu}$  and covariance  $\check{C}$  are calculated using Eq. 2 and Eq. 3 replacing the points  $p_{c_i}$  with their associated colour values  $\check{p}_{c_i} = [C^1, C^2, C^3]$ .

We use the *RGB* colour space in our difference detection system. To obtain some degree of invariance against changing illumination, *RGB* values are converted to the *YUV* colour space, the intensity *Y* is set to a constant value of 0.5, and then the *YUV* values are converted back to *RGB*.

## C. Adaptive Cell Splitting with Colour

Colour cell splits are generally performed in the same way as the cell splits based on Cartesian coordinates described above. However, in the colour space the criterion for a split is only dependent on the highest eigenvalue  $\check{\lambda}^{max}$  of the covariance matrix  $\check{C}$ . A colour split is carried out if  $\check{\lambda}^{max}$  is larger than a predefined threshold. This threshold was set to 10 in this paper referring to *RGB* values in between 0 and 255. Compared to the volume measure in Eq. 4, a modified criterion is used since lines and planes in colour space do not generally correspond to consistent structures in the environment.

## V. DIFFERENCE PROBABILITY COMPUTATION

### A. Spatial Difference Probability

A probabilistic value of the point  $p$  being different from the reference model is computed using the 3D-NDT representation of the reference model. First, the cell  $c$  is determined that contains the point  $p$ . Using the mean  $\mu_c$  and covariance  $C_c$  of this cell the spatial difference probability is then calculated as

$$p_{diff}(p) \propto e^{-(p-\mu_c)^T C_c^{-1} (p-\mu_c)}. \quad (6)$$

If adaptive cell splitting is used and therefore each grid cell can contain multiple mean and covariance values, the difference probability  $p_{diff}$  is calculated for all sub-cells using Equation 6 and the lowest probability is assigned to point  $p$ .

Note that the difference probability will depend on the structure of the point cloud used to generate the 3D-NDT representation of the reference model. In the case of planar regions, for example, the covariance matrix in Equation 6 will emphasise differences orthogonal to the planar structure so that even very small deviations can be detected there.



Fig. 5. Our mobile robot platform “Tjorven” equipped with the sensors used in this paper: the SICK LMS 200 laser range scanner and a colour CCD camera both mounted on an Amtec pan tilt unit.

### B. Colour Difference Probability

In the same way as the spatial difference probability, the colour difference probability is calculated as

$$p_{colordiff}(\check{p}) \propto e^{-(\check{p}-\check{\mu}_c)^T \check{C}_c^{-1} (\check{p}-\check{\mu}_c)}. \quad (7)$$

As described in the previous section, the difference probability will vary depending on the colour distribution of the selected cell, giving higher probabilities to colour changes in regions of very uniform colour distribution.

## VI. VALIDATION EXPERIMENT

### A. Hardware

For the experiments presented in this paper we used the ActivMedia P3-AT robot “Tjorven” shown in Fig. 5, equipped with a 2D laser range scanner (SICK LMS 200) and a 1-MegaPixel (1280x960) colour CCD camera. The CCD camera and the laser scanner are both mounted on a pan-tilt unit from Amtec with a displacement between the optical axes of approx 0.2 m. The angular resolution of the laser scanner was set to 0.25 degrees.

### B. Data Collection

For each pose, 3D range and image data are collected as follows. First, three sweeps are carried out with the laser scanner at -60, 0 and 60 degrees relative to the robot orientation (horizontally). During each of these sweeps, the tilt of the laser scanner is continuously shifted from -40 degrees (looking up) to 30 degrees (looking down). After the three range scan sweeps, seven camera images are recorded at -90, -60, -30, 0, 30, 60, and 90 degrees relative to the robot orientation (horizontally) and at a fixed tilt angle of -5 degrees (looking up). The full data set acquired at a single scan pose is shown on Fig. 6.

### C. Experiment

To evaluate the suggested difference detection system, a reference model of an indoor lab environment was created from the data recorded at 22 robot poses (reference data), i.e. from 66 laser scanner sweeps and 154 camera images. The reference model is shown in Fig. 7. Then, controlled changes (described in the Results section VI-D) were introduced and

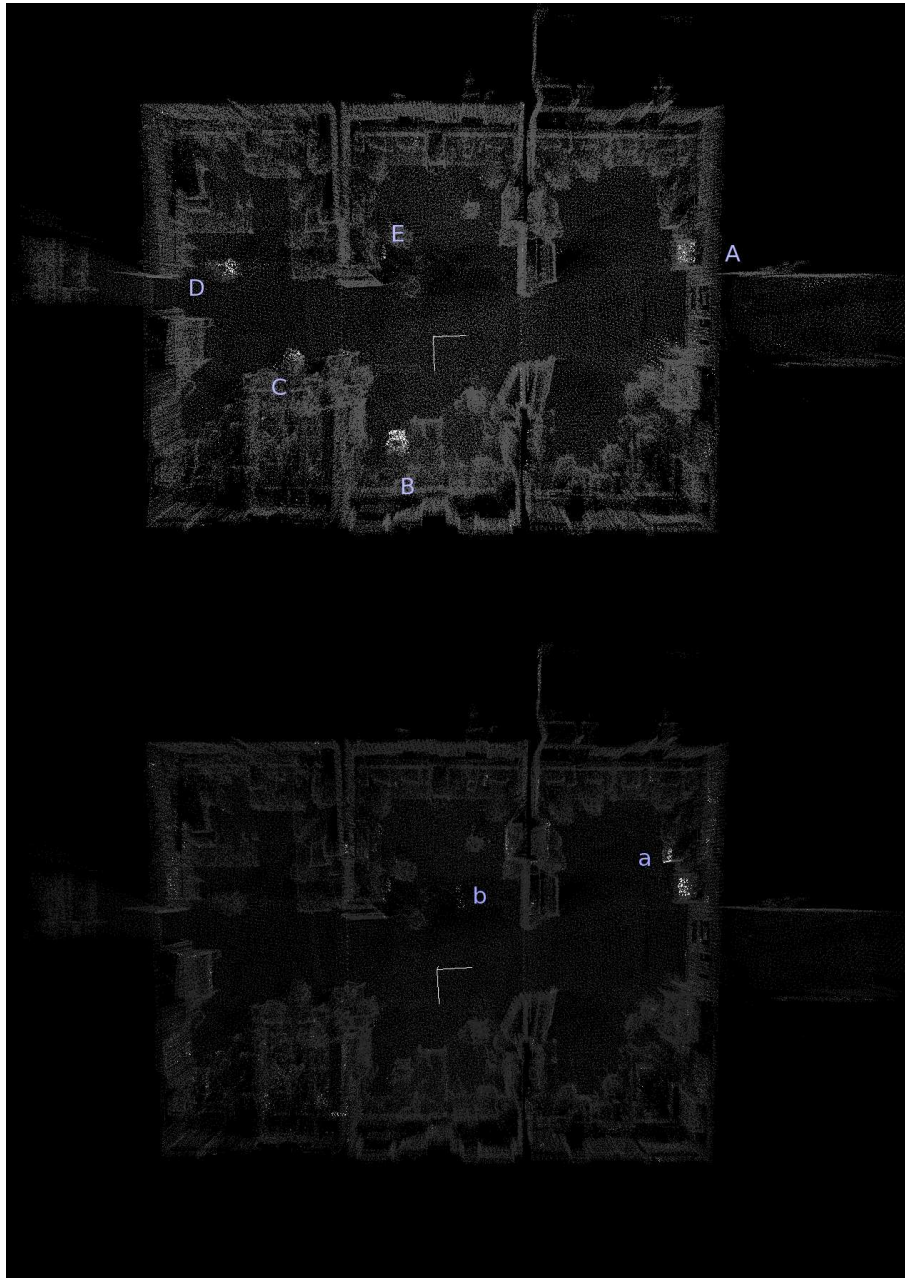


Fig. 8. Difference probability as found in 24 scans compared to the reference model shown in Fig. 7. Brighter regions indicate a higher difference probability. The scans are registered for better visualisation. Top: Difference probability computed from range values only using Eq. 6. Bottom Difference probability computed from colour information using Eq. 7.

a new data set was recorded at 24 different poses (difference detection data). The two data sets overlap each other and the difference detection data were treated independently from each other, i.e. an a priori unknown position was assumed for the 24 difference detection data sets. Note that no global optimization is carried out when creating the reference model. With the given set-up we found that using sequential registration the accumulated error was only a few centimeters between the first and last data set. In order to cope with larger environments, global optimization has to be integrated into the system, which is indeed in the pipe for future work.

#### D. Results

Fig. 8 shows the result of the evaluation experiment described in Section VI-C: the point cloud of the combined difference detection scans (registered for better visualization) shaded according to the computed difference probability. Brighter regions indicate a higher probability of changes in the environment. Fig. 8, top shows the difference probability obtained from spatial data only, i.e. applying Eq. 6. Fig. 8, bottom shows the difference probability obtained from colour data only, i.e. using Eq. 7.

All the changes to the reference model can be found in

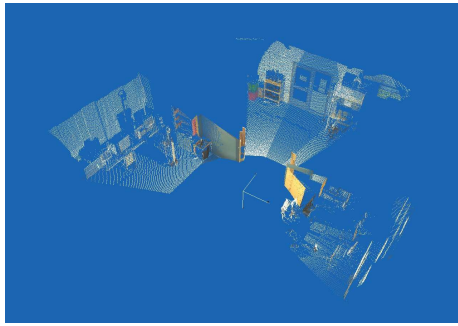


Fig. 6. Full data set acquired for a single scan pose comprising three sweeps with the laser scanner fused with colour information from seven camera images.



Fig. 7. The reference model created in our evaluation experiment consisting of  $3 \times 22$  registered scans and the corresponding colours from  $7 \times 22$  camera images.

the difference probability point clouds. The changes are indicated with an alphabetic character in Fig. 8: a sliding door that was opened (A), three chairs that were moved (B,C,D), and a relatively small box of approx.  $0.15 \times 0.25 \times 0.4 \text{ m}^3$  that was re-positioned (E). Further changes are not detectable using range data only but can be found when using colour data. These changes are two equally sized boxes (approx.  $0.4 \times 0.4 \times 0.5 \text{ m}^3$ ) that were swapped (a) and a coloured paper stripe (approx.  $0.1 \times 0.7 \text{ m}^2$ ) that was fixed to the floor (b).

## VII. CONCLUSION AND FUTURE WORK

In this paper we present a system for autonomous change detection with a security patrol robot. The method uses vision and 3D range data to build a reference model of the environment and detect changes with respect to this model. The approach was verified in a real world experiment in an indoor lab environment, demonstrating that the suggested system is able register 3D scans and to detect changes in spatial data and also to detect changes that occur in colour space and are not observable using range values only. Apart from the description of this system for autonomous difference detection, the particular contribution of this paper is the introduction of novel methods for vision-aided scan registration, which essentially works without an initial pose estimate, and using the 3D-NDT representation with adaptive cell size to efficiently represent both the spatial and colour aspects of the reference model.

Our future work will include a more thorough evaluation of the capabilities and limitations of the suggested system. Next, changes are currently detected in the coloured point clouds, which means that the resolution of the 3D range sensor is used. The higher resolution of the camera images is so far only exploited for scan registration. It is therefore another direction of our current research to extend the change detection system such that it uses “super-resolution” depth information obtained from a low-resolution 3D range sensor and a color image also for change detection. Finally, we want to investigate the possibility to continuously create the reference model in an autonomous manner, i.e. to represent the reference model as a dynamic map which would allow detecting changes at different time scales.

## REFERENCES

- [1] “Securitas Inc.” <http://www.securitas.com>.
- [2] J. N. K. Liu, M. Wang, and B. Feng, “ibotguard: an internet-based intelligent robot security system using invariant face recognition against intruder,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, vol. 35, no. 1, pp. 97–105, 2005.
- [3] M. Seeman, M. Broxvall, A. Saffiotti, and P. Wide, “An autonomous spherical robot for security tasks,” in *IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety (CIHSPS)*, 2006.
- [4] M. A. Goodrich, D. R. O. Jr., J. W. Crandall, and T. J. Palmer, “Experiments in adjustable autonomy,” in *Proceedings of IJCAI Workshop on Autonomy, Delegation and Control: Interacting with Intelligent Agents*, 2001.
- [5] P. Biber and T. Duckett, “Dynamic Maps for Long-Term Operation of Mobile Service Robots,” in *Proceedings of Robotics: Science and Systems I*, Cambridge, MA, USA, June 8–11 2005.
- [6] K. Ishikawa, J. ichi Takiguchi, and M. Hatayama, “Parking-vehicles recognition using spatial temporal data (a study of mobile robot surveillance system using spatial temporal gis part 2),” in *IEEE International Workshop on Safety, Security and Rescue Robotics (SSRR)*, 2005.
- [7] H. Andreasson, R. Triebel, and A. Lilienthal, “Vision-based interpolation of 3d laser scans,” in *Proceedings of the 2006 IEEE International Conference on Autonomous Robots and Agents (ICARA 2006)*, Palmerston North, New Zealand: IEEE, 2006.
- [8] D. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, November 2004.
- [9] P. M. Newman, D. M. Cole, and K. L. Ho, “Outdoor SLAM using visual appearance and laser ranging,” in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 2006, pp. 1180–1187.
- [10] J. Gonzalez-Barbosa and S. Lacroix, “Rover localization in natural environments by indexing panoramic images,” in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 2002, pp. 1365–1370.
- [11] P. J. Besl and N. D. McKay, “A Method for Registration of 3-D Shapes,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.
- [12] Y. Chen and G. Medioni, “Object Modelling by Registration of Multiple Range Images,” *Image and Vision Computing*, vol. 10, no. 3, pp. 145–155, 1992.
- [13] K. Mikolajczyk and C. Schmid, “A Performance Evaluation of Local Descriptors,” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005. [Online]. Available: <http://lear.inrialpes.fr/pubs/2005/MS05>
- [14] K. Arun, T. Huang, and S. Blostein, “Least-squares fitting of two 3-d point sets,” *Trans. PAMI*, vol. 9, no. 5, 1987.
- [15] P. Biber and W. Straßer, “The normal distributions transform: A new approach to laser scan matching,” in *IEEE/RJS International Conference on Intelligent Robots and Systems (IROS)*, 2003.
- [16] M. Magnusson and T. Duckett, “A comparison of 3d registration algorithms for autonomous underground mining vehicles,” in *Proceedings of the European Conference on Mobile Robotics (ECMR 2005)*, 2005.
- [17] M. Magnusson, T. Duckett, and A. Lilienthal, “3D Scan Registration for Autonomous Mining Vehicles,” *Journal of Field Robotics*, p. to appear, 2007.