Change detection on UGV patrols with respect to a reference tour using VIS imagery

Thomas Müller
Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB)
Fraunhoferstr. 1, 76131 Karlsruhe, Germany

ABSTRACT

Autonomous driving robots (UGVs, Unmanned Ground Vehicles) equipped with visual-optical (VIS) cameras offer a high potential to automatically detect suspicious occurrences and dangerous or threatening situations on patrol. In order to explore this potential, the scene of interest is recorded first on a reference tour representing the ‘everything okay’ situation. On further patrols changes are detected with respect to the reference in a two step processing scheme. In the first step, an image retrieval is done to find the reference images that are closest to the current camera image on patrol. This is done efficiently based on precalculated image-to-image registrations of the reference by optimizing image overlap in a local reference search (after a global search when that is needed). In the second step, a robust spatio-temporal change detection is performed that widely compensates 3-D parallax according to variations of the camera position. Various results document the performance of the presented approach.

Keywords: Change detection, detection of dangerous situations, image retrieval, image registration, 3-D parallax, UGV (Unmanned Ground Vehicle), Counter-IED, Security and Safety.

1. INTRODUCTION

In a variety of security and safety applications, autonomous driving robots, i.e. Unmanned Ground Vehicles (UGVs), that are equipped with visual-optical (VIS) cameras can perform valuable recognition tasks and inform about unusual situations or potentially critical incidents. For example, after a catastrophe UGVs can inspect all areas of interest on cyclic patrols and detect potential dangers by finding areas where e.g. building parts have moved or collapsed between patrols. Furthermore, danger sources can be identified by detecting their observable effects in a scene.

As a second example, the protection of field camps or real estates can be mentioned here. If some manipulation or sabotage has been done between robot patrols, i.e. objects have been placed somewhere (like traps, barriers, bombs/Improvised Explosive Devices (IEDs), or ladders), moved (e.g. windows, doors, vehicles), or damaged (e.g. a fence or a door) all that can potentially be detected by an autonomous UGV by detecting corresponding scene changes.

All mentioned cases have in common that scene changes are happening between one robot patrol and a subsequent one which can potentially be seen by an on-board sensor. The focus of this paper is to detect such scene changes based on image sequences stemming from the VIS camera on-board the UGV. Of course, this could alternatively be done also with a 3-D laser sensor, which would mean to emphasize more the geometrical aspect whereas the focus with a VIS camera is more on color and appearance. So, both sensors have advantages and disadvantages. The decision in this paper is to use VIS imagery because the appearance aspect seems to be more important in some concrete practical applications in focus (like the detection of buried traps or IEDs by means of appearance changes of the ground, or the detection of victims after a catastrophe).

This paper is organized as follows: In Section 2 some related work is discussed. Afterwards, Section 3 gives an overview over the developed overall system and its components, which are described then in detail in Sections 4 to 7. After this, Section 8 presents some achieved results. The paper concludes with a summary and outlook in Section 9.

Further author information:
Thomas Müller: E-mail: Thomas.Mueller@iosb.fraunhofer.de, Telephone: +49(0)721/6091 458.
2. RELATED WORK

Most papers in the field of change detection for UGVs or for moving cameras are dealing with and, therefore, are restricted to the detection of moving objects in the scene\textsuperscript{3–5} or change detection is done mainly for the purpose of collision or damage avoiding on patrol\textsuperscript{9–11}.

Changes in a static environment are detected with 3-D LIDAR sensors\textsuperscript{9, 12} or with VIS sensors based on a 3-D reconstruction of the scene with one camera,\textsuperscript{13} scene segmentation,\textsuperscript{10, 11} image segmentation with Superpixels,\textsuperscript{14} or motion blob detection that is combined with Split Gaussian Models.\textsuperscript{15} Further contributions do an automatic image registration based on SIFT-Features\textsuperscript{16} followed by a temporal image differencing in indoor scenes,\textsuperscript{17} or a processing chain is performed consisting of image retrieval, image differencing, post-processing using Markov Random Fields, local texture comparison, and change tracking.\textsuperscript{18} Similar to the applications discussed here, some papers perform a change detection using UAVs (Unmanned Aerial Vehicles). For example, the camera images are registered using SIFT-features\textsuperscript{16} and processed with a multi-primitive segmentation procedure,\textsuperscript{19} or an image registration is used to calculate image mosaics on which image changes are processed then.\textsuperscript{20} It should be mentioned that such UAV works rely on orthophotos and, therefore, cannot be used just like that also on UGVs due to the very different camera perspective and camera movement through the scene leading to highly different conditions.

3. SYSTEM ARCHITECTURE

As stated in the introduction, changes have to be detected between two patrols of the UGV (see utilized UGV in Figure 1). The first patrol serves as reference. The VIS image sequence on the reference patrol is recorded, saved to disc and processed afterwards in order to precalculate all necessary informations that will be needed later. This preprocessing step is explained in detail in Section 4.

Then, on a later patrol the live image data has to be compared with the reference image data online to detect scene changes. Modern SLAM and navigation methods\textsuperscript{1, 21} on the UGV try to reproduce the robot trajectories on different patrols with a good repeat accuracy. But, of course, since intelligent path planning has to be dynamic in order to react properly on environmental changes or obstacles, differences between two patrols are unavoidable. So, between patrols there will be differences in the robot position and/or orientation and, therefore, also in the camera pose that have to be taken into account. Such differences can be small with little differences between corresponding images. But greater differences can occur, too, that lead to big displacements between corresponding images – like 100 or 300 pixel to give an example. And additionally, even greater differences can be produced by the path planning algorithm for a period of time so that on the current patrol areas are seen by the camera that have not been recorded previously, for example when the path planning process reacts to slightly different environmental conditions with significant differences in the calculated trajectory. Furthermore, path planning can also produce different driving speeds or different durations of stay at dedicated path positions or in particular environmental situations.
In order to handle big differences between images or patrols, an image retrieval is needed in order to find corresponding images in the reference image data base that presumably match best to the current camera image on patrol. In this procedure not only the best image is calculated but the \( n \) best reference images so that the change detection step can perform a combined calculation over all \( n \) images in order to compensate for image differences in these images and the current camera image. 'Best reference images' means with respect to a novel, simple and therefore efficient criterion considering image overlap and matching (cf. Section 5). In order to compensate for time differences and spatial differences between reference and current patrol, a local image retrieval searches for the best reference images in an appropriate neighborhood that is defined in a spatial space given by the criterion. Usually, this local search can just start using the images found for the previous image. If this fails, for example because an area is observed that is not depicted in the reference data base, a global image retrieval procedure searches the reference data base until a currently recorded area can be found again in some reference image. As soon as a reference image can be found again, the local image retrieval is processed to find also the \( n \) best image candidates afterwards. In Sections 5 and 6 the global and local image retrieval are described in detail, respectively. After the reference images have been found, a robust change detection is performed which compensates for different camera positions and view angles. This final processing step is explained in Section 7.

In order to handle small differences between images, a robust image-to-image registration\(^{26,27}\) is applied to warp the image content of one image onto another one so that image contents at same pixel positions are comparable with each other. Additionally, image registrations between successive images are calculated as a base to indirectly estimate some kind of movement speed on patrol in order to adjust different robot speeds and durations of stay between reference and current patrol. Furthermore, miscellaneous preceding experiments have shown our image registration as sufficiently robust to also check to a large extent, if two images show nearly the same area or not on the same patrol or between different patrols. Therefore, the image registration is applied as verification procedure if two images match or not, too. But it has to be considered here that this works properly only if the overlap between two considered images is sufficient. If the overlap is too small, images that show the same area will be classified as unmatchable, though. Therefore, the global image retrieval has to check all reference images with an inter-distance below some threshold for critical image overlap on the one hand. On the other hand, it is sufficient to check only reference images with an inter-distance above some smaller threshold \( \delta \) in order to avoid lots of time-consuming registrations with unnecessary similar images. The above mentioned speed estimation is used to efficiently calculate all needed distances between images in the reference image sequence. In this manner, only a minimum of images is selected in the reference data base to be matched with the current image on patrol in order to minimize calculation time.

If an area to be observed by the UGV is very large, the presented approach can easily be combined with GPS data on-board the UGV in order to select the appropriate sector in the reference data base that has to be used by the image retrieval process.

**4. REFERENCE DATA PROCESSING**

After an image sequence has been recorded on a patrol that shall serve as reference for one or more further patrols, the images are saved to disc and processed in order to preprocess all informations that are needed in the image retrieval and change detection processes later. In this procedure the image-to-image registration is computed between every image \( i \) and the previous image \( i - 1 \) (for \( i = 2, \ldots, N \) with \( N \) denoting the length of the recorded image sequence). In case of success, i.e. both images can be matched, a success flag \( s_{i-1,i} \) is set to 1 and a homography \( R_{i-1,i} : R^2 \rightarrow R^2 \) is computed with which image coordinates \((x, y)^T\) of image \( i \) can be warped onto the image coordinates of image \( i - 1 \). In case of failure, for example if both images show different areas or things for some reason that cannot be matched, the success flag is set to \( 0 \).

Additionally, every image \( i \) of the second half of the image sequence is tried to be registrated with the first image of the sequence in order to detect a cyclic patrol. This is only done for the second half of the sequence to avoid false cycle detections at the beginning of the sequence to the beginning itself. If there are images \( i \) that can be matched (i.e. with \( s_{1,i} = 1 \) along with a computed homography \( R_{1,i} \)), then image number \( j \) is computed that leads to the minimal displacement of the image center \((w/2, h/2)^T\) (\( w \) and \( h \) denoting the pixel width and
If a cycle was found, \( N \) is set to \( j \) (to cut the reference sequence off here, because the path planning just repeats the patrol at this point then) and \( s_{0,1} := 1 \), \( R_{0,1} := R^{-1}_{R,j} \) to close the established homography chain. Finally, the following data is saved to file: \( N \), a remark if the patrol is cyclic or not, and \( s_{i-1,i}, R_{i-1,i} \) for \( i = 1, \ldots, N \). Inverse homographies \( R_{i-1,i} = R^{-1}_{R,i-1,i} \) are required later, too, but it is not needed to save them in the file, since they can just be computed while reading the file at the initialization of the image retrieval.

5. GLOBAL IMAGE RETRIEVAL

Before the change detection step, an image retrieval procedure has to find images in the reference data base that match with the current image on patrol. Most of the time it is sufficient to perform a local image retrieval beginning with the reference image that was found for the previous image on the current patrol, because successive images on reference and current patrol usually exhibit only small displacements in between. Local image retrieval means that the search is performed around the matching reference image to a reasonable extent. But if the local search cannot be applied or fails for some reason (cf. Section 3), a global image retrieval has to ensure that matching images in the reference are found as soon as it is possible again. Generally, this search has to include the whole reference data base, since no made assumptions should exclude any regions that were seen by the camera on the reference patrol. This is important especially in cases where the UGV has planned a trajectory on patrol that is very different from the reference patrol. Because with increasing time without a match in the reference data base, the potential UGV pose gets more and more uncertain and a ‘reappearing’ with respect to the reference can potentially happen at any arbitrary point in the reference image sequence again.

The image retrieval step starts with trying to calculate the image registration \( T_{R,P} \) between the current image on patrol and the best reference image that has been found for the previous image (or the first image of the reference sequence, if the current patrol has just started with its first image). If this fails, the last reference image number is predicted temporally according to the measured computation time of the overall algorithm for the previous image in order to compensate a potential deviation just due to that delay and, additionally, it is tried to calculate the image registration \( T_{R,P} \) between the current image on patrol and this predicted reference image, too. If the current image on patrol can be matched in this way with a reference image, the algorithm continues with the local image retrieval (cf. following Section 6). Otherwise, the global image retrieval is processed first in order to find a reference image and determine the respective \( T_{R,P} \) along with it.

The global image retrieval procedure starts with the image number \( i \) of the last reference image that could be matched with some previous image on the current patrol in the history before (or with the first image \( i = 1 \) of the reference sequence, if the current patrol has just started with its first image). Iteratively, a successive warping of the image center \( (w/2, h/2)^T \) is used as an estimation of moving speed over the image sequence in order to perform a simultaneous forward and backward search for a relatively small number of images that have to be matched with the current image on patrol. I.e. iteratively for \( k = 1, \ldots, N/2 \) the minimal \( k \) is determined that leads to

\[
\begin{align*}
& s_{i+k,i+k-1} = 0 \lor \left\| R_{i+k,i+k-1} \circ \ldots \circ R_{i+1,i} \left( \begin{pmatrix} w/2 \\ h/2 \end{pmatrix} \right) - \begin{pmatrix} w/2 \\ h/2 \end{pmatrix} \right\| \geq \delta \\
& \text{(forward search)} \\
& s_{i-k,i-k+1} = 0 \lor \left\| R_{i-k,i-k+1} \circ \ldots \circ R_{i+1,i} \left( \begin{pmatrix} w/2 \\ h/2 \end{pmatrix} \right) - \begin{pmatrix} w/2 \\ h/2 \end{pmatrix} \right\| \geq \delta \\
& \text{(backward search)}
\end{align*}
\]
at the end of Section 3 (we use $\delta = 150$ pixel for images of size $720 \times 576$ pixel to give a concrete value here).

All image numbers and, therefore, indices of $s_{i,*}$ and $R_{i,*}$ are understood modulo $N$ in this paper.

If $k$ is found in this described manner, reference image $i + k$ (forward) or $i - k$ (backward), respectively, is matched with the current image on patrol. In case of success, the global image retrieval is finished and the found reference image and the calculated homography $T_{R,P}$ are fed into the local image retrieval. Otherwise, the iteration with $k$ continues analogously, i.e. with a substituted transformation chain $R_{i,*} \circ \ldots \circ R_{i,*}$ beginning with the respective one element $R_{i,*}$ for the next $k$ in the forward or backward case, respectively, so that the euclidean measurement of distances in the formulas starts in the further iterations with respect to the reference image that has been matched.

The described procedure selects a relatively small subset of reference images that have to be matched, skips short and long durations of stay and therefore compensates different durations of stay between patrols. Only reference images showing sufficient different areas are taken into account in order to save computing time. This is done as simultaneous forward and backward search because a next hit can initially be assumed as more probable around the last hit than far away from it.

6. LOCAL IMAGE RETRIEVAL

The local image retrieval receives as input a reference image with number $i$ that could be matched before with the current image on patrol and the calculated homography $T_{R,P}$ between the images. Starting from the given reference image, $n$ images are searched in the reference data base that are closest to the current image with respect to a certain distance measurement. Initially, a similar distance measurement was used here as in the global image retrieval based on a successive warping of the image center $(w/2, h/2)^T$. But experiments have shown that the use of more than one image point produces better results at the end – obviously because image distortions are also taken into account. So, a different measurement is used in the local image retrieval. Best results were obtained with three or four image points. In the experiments in this paper four image points were used: $(x_1, y_1)^T := \left(\frac{w}{8}, \frac{h}{8}\right)^T$, $(x_2, y_2)^T := \left(\frac{w}{8}, \frac{7h}{8}\right)^T$, $(x_3, y_3)^T := \left(\frac{w}{8}, \frac{h}{8}\right)^T$ and $(x_4, y_4)^T := \left(\frac{w}{8}, \frac{7h}{8}\right)^T$.

Iteratively, via successive warping of these image points (first in forward and then in backward direction, not necessarily simultaneously as in the global search) an efficient estimation is made about image overlap and matching between the current image on patrol and all reference images in a (roughly estimated) near range. I.e. iteratively for $k = 1, 2, 3, \ldots$ all distance measurements in forward direction

\[ d_k^{(f)} := \max_{j=1,2,3,4} \left\| R_{i+k,i+k-1} \circ \ldots \circ R_{i+1,i} \circ T_{R,P}(\begin{bmatrix} x_j \\ y_j \end{bmatrix}) - \begin{bmatrix} x_j \\ y_j \end{bmatrix} \right\| \]  

are calculated until the first $k$ with $s_{i+k,i+k-1} = 0$ or $d_k^{(f)} > \gamma$ with a big threshold $\gamma$ (which in fact has to be relatively big because, on the one hand, the reference image $i$ can potentially be just a very rough initial matching candidate so that the iterative process needs to overcome big displacements first before smaller ones can arise then. On the other hand, local (potentially relatively big) maxima are overrun in the iterations so that important images afterwards with small values can be collected, too. The bigness of $\gamma$ is no problem since the calculations that have to be made are not very time consuming. $\gamma$ limits the number of calculated $d_k^{(f)}$ by roughly cutting off far range in the reference data base. We use $\gamma = 500$ pixel). The chain of homographies in the formula is used as an estimation of the homography between the current image on patrol and reference image $i + k$. Due to potential drift over time and accumulating errors the resulting homography has to be recomputed in the change detection step where needed. The calculations are done in an analogous manner also in backward direction

\[ d_k^{(b)} := \max_{j=1,2,3,4} \left\| R_{i-k,i-k+1} \circ \ldots \circ R_{i-1,i} \circ T_{R,P}(\begin{bmatrix} x_j \\ y_j \end{bmatrix}) - \begin{bmatrix} x_j \\ y_j \end{bmatrix} \right\| \]  

until the first $k$ with $s_{i-k,i-k+1} = 0$ or $d_k^{(b)} > \gamma$. Finally, the $n$ lowest values are determined in the series of all $d_k^{(f)}$ and $d_k^{(b)}$ leading to $n$ image numbers of reference images that are presumably matching best with the current image on patrol.
7. CHANGE DETECTION STEP

In the change detection step, the \( n \) images of the reference data base that were found by the local image retrieval are fed into an algorithm that was originally developed for robust temporal change detection in the field of MTI (Moving Target Indication). This algorithm\(^{26}\) could be transferred and adapted to the situation given here.

First, all \( n \) reference images are registred with the current image on patrol and warped so that all structures match spatially with the current image. Then, \( n \) specific difference images are calculated between the warped reference images and the current image with respect to the local image structure. In this procedure, for each pixel in the current image the local color dynamic is registered by determining the minimal and maximal intensity in the direct pixel neighborhood of size 4 so that an intensity interval can be defined for the pixel which models the local intensity dynamic. The difference of a warped reference image and the current image at a pixel is then calculated by comparing the reference pixel with the dynamic interval of the current image. If the pixel is likely to be part of the dynamic (i.e. the pixel intensity is an element of the interval) the resulting difference is 0. Otherwise, the difference to the nearer interval border is taken as difference result.

In order to further increase robustness and to compensate irrelevant image differences and particularly 3-D parallax of the depicted scene, the resulting \( n \) difference images are fused afterwards in a generalized AND operation in order to extract the changes that all difference images have in common. The final result is overlayed to the current image on patrol as user feedback.

Parameter \( n \) can be used as tuning parameter with which the presented change detection algorithm can be adapted either more to higher scene complexity or to a higher change detection sensitivity. Lower values lead to an increased sensitivity when little scene changes have to be detected, but small values can only be used for scenes with low 3-D parallax or image noise (like \( n = 4 \)). If there is much 3-D parallax, \( n \) has to be increased to compensate that, but simultaneously the detection sensitivity decreases. In the results presented in Section 8, a value of \( n = 20 \) was used because of much 3-D parallax (e.g. due to trees and buildings near the UGV path).

8. EXPERIMENTAL RESULTS

Figures 2 to 5 present some obtained results in different situations. The figures show the current image at patrol, the reference image \( k \) with lowest \( d_k^{(s)} \) found by the local image retrieval, the change detection result image (including also a scaled, small version of reference image \( k \) overlayed in the upper right corner as an additional feedback for the system user at runtime) and a magnification of the detections. The change detection results are marked with overlayed bright color (yellow in the electronic paper version and a kind of light grey in a black and white paper print). The results document the capability and benefit of the presented approach in different situations and applications.

In order to save computation time, all image registrations in the image retrieval and change detection steps are calculated on images that were downscaled before to half of the original resolution. Finally, the obtained change detection results are upscaled before overlaying them to the current image on patrol. The overall system runs most of the time (i.e. when no global image retrieval is needed) with a framerate between 1.7 and 2.5 Hz for camera images of 720 \( \times \) 576 pixel in size on an Intel Xeon PC with 3.60 GHz. If the global image retrieval has to be processed, overall calculation times between 0.5 and 2 seconds per image were measured in the experiments processing reference data sets containing up to 15,000 images.

9. CONCLUSIONS AND FUTURE WORK

A change detection system has been presented for UGVs on patrol based on VIS camera images. Changes are detected between two patrols of the robot. After a preprocessing of the first patrol, changes on a current patrol are detected using a robust change detection method after an efficient image retrieval step has collected appropriate images from the first patrol based on a novel, simple and fast measurement estimating moving speed in the image. The depicted results document the capability and benefit in different situations and applications.

In situations with much 3-D parallax, the presented method offers the possibility to compensate negative effects coming along with that to some degree. Future work should now deal with expanding this feature so that even more 3-D parallax can be handled while even smaller details can be detected. In order to lower the calculation time of the global search, the method could be merged with a generic image retrieval approach.\(^{28}\)
Figure 2. Upper left: Current image on patrol. Upper right: Reference image $k$ with lowest $d_k^{(*)}$ found by the local image retrieval. Lower row: Detection of an object positioned near the road that could be a potential danger source (marked in bright color; left) and magnification of the found region (right).
Figure 3. Upper left: Current image on patrol. Upper right: Reference image $k$ with lowest $d_k^{(*)}$ found by the local image retrieval. Lower row: Detection that two cars behind another car have left the car park (marked in bright color; left) and magnification of the found region (right).
Figure 4. Upper left: Current image on patrol. Upper right: Reference image $k$ with lowest $d_k^{(c)}$ found by the local image retrieval. Lower row: Detection that a person has appeared and a car has left the car park (marked in bright color; left) and magnification of the found region (right).
Figure 5. Upper left: Current image on patrol. Upper right: Reference image $k$ with lowest $d^*_k$ found by the local image retrieval. Lower row: Detection of changes in debris on the left in near UGV distance and detection of a collapse of a building part at farer distance near a car (marked in bright color; left) and magnification of the found region (right).
REFERENCES


