

Fusion of Aerial Images and Sensor Data from a Ground Vehicle for Improved Semantic Mapping

Martin Persson^{*a}, Tom Duckett^{**}, and Achim Lilienthal^{*}

*Centre of Applied Autonomous Sensor Systems
Department of Technology
Örebro University, Sweden
martin.persson@tech.oru.se,
achim@lilienthals.de

**Department of Computing and Informatics
University of Lincoln
Lincoln, UK
tduckett@lincoln.ac.uk

Abstract— This paper investigates the use of semantic information to link ground-level occupancy maps and aerial images. A ground-level semantic map is obtained by a mobile robot equipped with an omnidirectional camera, differential GPS and a laser range finder. The mobile robot uses a virtual sensor for building detection (based on omnidirectional images) to compute the ground-level semantic map, which indicates the probability of the cells being occupied by the wall of a building. These wall estimates from a ground perspective are then matched with edges detected in an aerial image. The result is used to direct a region- and boundary-based segmentation algorithm for building detection in the aerial image. This approach addresses two difficulties simultaneously: 1) the range limitation of mobile robot sensors and 2) the difficulty of detecting buildings in monocular aerial images. With the suggested method building outlines can be detected faster than the mobile robot can explore the area by itself, giving the robot an ability to “see” around corners. At the same time, the approach can compensate for the absence of elevation data in segmentation of aerial images. Our experiments demonstrate that ground-level semantic information (wall estimates) allows to focus the segmentation of the aerial image to find buildings and produce a ground-level semantic map that covers a larger area than can be built using the onboard sensors.

I. INTRODUCTION

A mobile robot has a limited view of its environment. Mapping of the operational area is one way of enhancing this view for visited locations. In this paper we explore the possibility to use information extracted from aerial images to further improve the mapping process. Semantic information (classification of buildings versus non-buildings) is used as the link between the ground level information and the aerial image. The method speeds up exploration or planning in areas unknown to the robot.

Colour image segmentation is often used to extract information about buildings from an aerial image. However, it is hard to perform automatic detection of buildings in monocular aerial images without elevation information. Buildings can not easily be separated from other man-made structures such as driveways, tennis courts, etc. due to the resemblance in colour and shape. We show that wall estimates found by a mobile robot can compensate for the absence of elevation data. In our previous work [19] wall estimates detected by a

mobile robot are matched with edges extracted from an aerial image. A virtual sensor¹ for building detection is used to identify parts of an occupancy map that belong to buildings (wall estimate). To determine potential matches we use geo-referenced aerial images and an absolute positioning system on board of the robot. The matched lines are then used in region- and boundary-based segmentation of the aerial image for detection of buildings.

In this paper, we extend the approach from [19]. The extension includes a global search for buildings in the aerial image and the introduction of a ground class. The purpose is to detect building outlines and driveable paths faster than the mobile robot can explore the area by itself. Using a method like this, the robot can estimate the size of found buildings and using the building outline it can “see” around one or several corners without actually visiting the area. The method does not assume a perfectly up-to-date aerial image, in the sense that buildings may be missing although they are present in the aerial image, and vice versa. It is therefore possible to use globally available² geo-referenced images.

A. Related Work

Overhead images in combination with ground vehicles have been used in a number of applications. Oh *et al.* [11] used map data to bias a robot motion model in a Bayesian filter to areas with higher probability of robot presence. Mobile robot trajectories are more likely to follow paths in the map and using the map priors, GPS position errors due to reflections from buildings were compensated. This work assumed that the probable paths were known in the map.

Pictorial information captured from a global perspective has been used for registration of sub-maps and subsequent loop-closing in SLAM [2].

Silver *et al.* [16] discuss registration of heterogeneous data (e.g. data recorded with different sampling density) from aerial surveys and the use of these data in classification of ground surface. Cost maps are produced that can be used in long range vehicle navigation. Scrapper *et al.* [15] used

¹A virtual sensor is understood as one or several physical sensors with a dedicated signal processing unit for recognition of real world concepts.

²E.g. Google Earth, Microsoft Virtual Earth, and satellite images from IKONOS and its successors.

^aSupported by The Swedish Defence Material Administration

heterogeneous data from, e.g., maps and aerial surveys to construct a world model with semantic labels. This model was compared with vehicle sensor views providing a fast scene interpretation.

For detection of man-made objects in aerial images, lines and edges together with elevation data are the features that are used most often. Building detection in single monocular aerial images is very hard without additional elevation data [18]. Mayer's survey [9] describes some existing systems for building detection and concludes that scale, context and 3D structure were the three most important features to consider for object extraction in aerial images. Fusion of SAR (Synthetic Aperture Radar) and aerial images has been employed for detection of building outlines [18]. The building location was established in the overhead SAR image, where walls from one side of buildings can be detected. The complete building outline was then found using edge detection in the aerial image. Parallel and perpendicular edges were considered and the method belongs to edge-only segmentation approaches.

Combination of edge and region information for segmentation of aerial images has been suggested in several publications. Mueller *et al.* [10] presented a method to detect agricultural fields in satellite images. First, the most relevant edges were detected. These were then used to guide both the smoothing of the image and the following segmentation in the form of region growing. Freixenet *et al.* [4] investigated different methods for integrating region- and boundary-based segmentation, and also claim that this combination is the best approach.

B. Outline and Overview

The presentation of our proposed system is divided into three main parts. The first part, Section II, concerns the estimation of walls by the mobile robot and edge detection in the aerial image. The wall estimates are extracted from a probabilistic semantic map. This map is basically an occupancy map that is labelled using a virtual sensor for building detection [12] mounted on the mobile robot. The second part describes the matching of wall estimates from the mobile robot with the edges found in the aerial image. This procedure is described in Section III. The third part presents the segmentation of an aerial image based on the matched lines. Section IV deals with a local segmentation to find buildings and Section V extends this to a global segmentation of the aerial image and also introduces the class driveable ground. Details of the mobile robot, the experiments performed and the obtained result are found in Section VI. Finally, the paper is concluded in Section VII and some suggestions for future work are given.

II. WALL ESTIMATION

A major problem for building detection in aerial images is to decide which of the edges in the aerial image correspond to building outlines. The idea of our approach, to increase the probability that a correct segmentation is performed, is to match wall estimates extracted from two perspectives. In this

section we describe the process of extracting wall candidates, first from the mobile robot's perspective and then from aerial images.

A. Wall Candidates from Ground Perspective

The wall candidates from the ground perspective are extracted from a semantic map acquired by a mobile robot. The semantic map we use is a probabilistic occupancy grid map augmented with labels for buildings and non-buildings [14]. The probabilistic semantic map is produced using an algorithm that fuses different sensor modalities. In this paper, a range sensor is used to build an occupancy map, which is converted into a probabilistic semantic map using the output of a virtual sensor for building detection based on an omnidirectional camera.

The algorithm consists of two parts. First, a local semantic map is built using the occupancy map and the output from the virtual sensor. The virtual sensor uses the AdaBoost algorithm [5] to train a classifier that classifies close range monocular grey scale images taken by the mobile robot as buildings or non-buildings. The method combines different types of features such as edge orientation, grey level clustering and corners into a system with high classification rate [12]. The classification by the virtual sensor is made for a whole image. However, the image may also contain parts that do not belong to the detected class, e.g., an image of a building might also include some vegetation such as a tree. Probabilities are assigned to the occupied cells that are within a sector representing the view of the virtual sensor. The size of the cell formations within the sector affects the probability values. Higher probabilities are given to larger parts of the view, assuming that larger parts are more likely to have caused the view's classification [14].

In the second step the local maps are used to update a global map using a Bayesian method. The result is a global semantic map that distinguishes between buildings and non-buildings. An example of a semantic map is given in Figure 1. From the global semantic map, lines representing probable building outlines are extracted. An example of extracted lines is given in Figure 2.

B. Wall Candidates in Aerial Images

Edges extracted from an aerial image are used as potential building outlines. We limit the wall candidates used for matching in Section III to straight lines extracted from a colour aerial image taken from a nadir view. We use an output fusion method for the colour edge detection. The edge detection is performed separately on the three RGB-components using Canny's edge detector [1]. The resulting edge image I_e is calculated by fusing the three binary images obtained for the three colour components with a logical OR-function. Finally a thinning operation is performed to remove points that occur when edges appear slightly shifted in the different components. For line extraction in I_e an implementation by Peter Kovcsi³ was used. The lines extracted from

³<http://www.csse.uwa.edu.au/~pk/Research/MatlabFns/>, University of Western Australia, Sep 2005

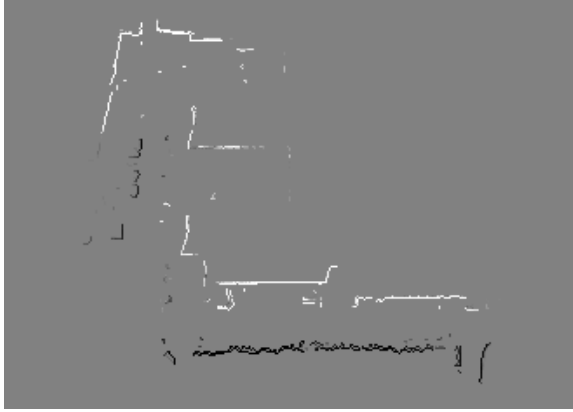


Fig. 1. An example of a semantic map where white lines denote high probability of walls and dark lines show outlines of non-building entities.

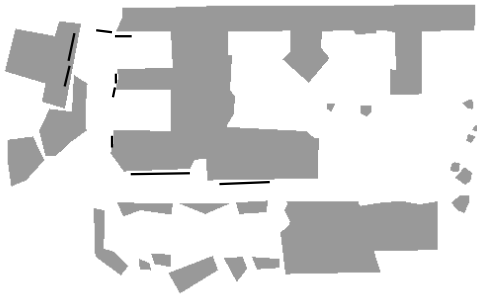


Fig. 2. Illustration of the wall estimates calculated from the semantic map (lines drawn in black). The grey areas illustrates building and nature objects (manually extracted from Fig. 3). The semantic map in Fig. 1 belongs to the upper left part of this figure.

the edges detected in the aerial image in Figure 3, are shown in Figure 4.

III. WALL MATCHING

The purpose of the wall matching step is to relate a wall estimate, obtained at ground-level with the mobile robot, to the edges detected in the aerial image. In both cases the line segments represent the wall estimates. We denote a wall estimate found by the mobile robot as L_g and the N lines representing the edges found in the aerial image by L_a^i with $i \in \{1, \dots, N\}$. Both line types are geo-referenced in the same Cartesian coordinate system.

The lines from both the aerial image and the semantic map may be erroneous, especially concerning the line endpoints, due to occlusion, errors in the semantic map, different sensor coverage, etc. We therefore need a metric for line-to-line distances that can handle partially occluded lines. We do not consider the length of the lines and restrict the line matching to the line directions and the distance between two points, one point on each line. The line matching calculations are performed in two sequential steps: 1) decide which points



Fig. 3. The trajectory of the mobile robot and the used aerial image.

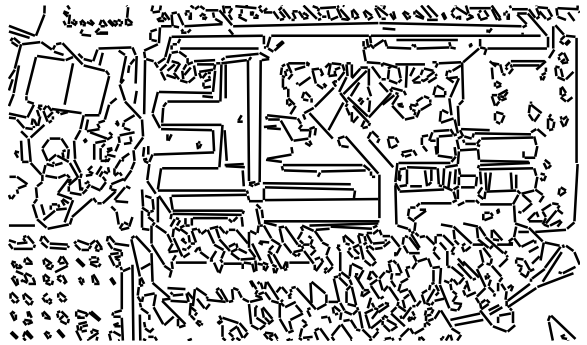


Fig. 4. The lines extracted from the edge version of the aerial image.

on the lines are to be matched, and 2) calculate a distance measure to find the best matches.

A. Finding the Closest Point

In this section we define which points on the lines are to be matched. For L_g we use the line midpoint, P_g . Due to the possible errors described above we assume that the point P_a on L_a^i that is closest to P_g is the best candidate to be used in our ‘line distance metric’.

To calculate P_a , let e_n be the orthogonal line to L_a^i that intersects L_g in P_g , see Figure 5. We denote the intersection between e_n and L_a^i as ϕ where $\phi = e_n \times L_a^i$ (using homogeneous coordinates). The intersection ϕ may be outside the line segment L_a^i , see right part of Figure 5. We therefore need to check if ϕ is within the endpoints and if it is set $P_a = \phi$. If ϕ is not within the endpoints, then P_a is set to the closest endpoint on L_a^i .

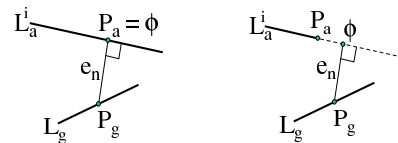


Fig. 5. The line L_g with its midpoint $P_g = (P_{gx}, P_{gy})$, the line L_a^i , and the normal to L_a^i , e_n . To the left, $P_g = \phi$ since ϕ is on L_a^i and to the right, P_g is the endpoint of L_a^i since ϕ is not on L_a^i .

B. Distance Measure

The calculation of a distance measure is inspired by [7], which describes geometric line matching in images for stereo matching. We have reduced the complexity in those calculations to have fewer parameters that need to be determined and to exclude the line lengths. Matching is performed using L_g 's midpoint P_g , the closest point P_a on L_a^i and the line directions, θ_g and θ_a . First, a difference vector is calculated as

$$\mathbf{r}_g = [P_{g_x} - P_{a_x}, P_{g_y} - P_{a_y}, \theta_g - \theta_a]^T. \quad (1)$$

Second, the similarity is measured as the Mahalanobis distance

$$\mathbf{d}_g = \mathbf{r}_g^T \mathbf{R}^{-1} \mathbf{r}_g \quad (2)$$

where the diagonal covariance matrix \mathbf{R} is defined as

$$\mathbf{R} = \begin{bmatrix} \sigma_{Rx}^2 & 0 & 0 \\ 0 & \sigma_{Ry}^2 & 0 \\ 0 & 0 & \sigma_{R\theta}^2 \end{bmatrix} \quad (3)$$

with σ_{Rx} , σ_{Ry} , and $\sigma_{R\theta}$ being the expected standard deviation of the errors between the ground-based and aerial-based wall estimates.

IV. LOCAL AERIAL IMAGE SEGMENTATION

This section describes how local segmentation of the colour aerial image is performed. Segmentation methods can be divided into two groups; discontinuity- and similarity-based [6]. In our case we combine the two groups by first performing an edge based segmentation for detection of closed areas and then colour segmentation based on a small training area to confirm the areas' homogeneity. The following is a short description of the sequence that is performed for each line L_g :

- 1) Sort the set of lines L_a based on \mathbf{d}_g from Equation 2 in increasing order and set $i = 0$.
- 2) Set $i = i + 1$.
- 3) Define a start area A_{start} on the side of L_a^i that is opposite to the robot (this will be in or closest to the unknown part of the occupancy grid map).
- 4) Check if A_{start} includes edge points (parts of edges in I_e). If yes, return to step 2.
- 5) Perform edge controlled segmentation.
- 6) Perform homogeneity test.

The segmentation based on L_g is stopped when a region has been found. Step 4 makes sure that the regions have a minimum width. Steps 5 and 6 are elaborated in the following paragraphs.

A. Edge Controlled Segmentation

Based on the edge image I_e constructed from the aerial image, we search for a closed area. Since there might be gaps in the edges bottlenecks need to be found [10]. We use morphological operations, with a 3×3 structuring element, to first dilate the interesting part of the edge image in order

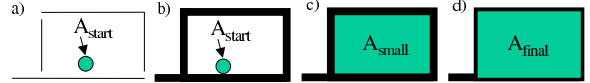


Fig. 6. Illustration of the edge-based algorithm. a) shows a small part of I_e and A_{start} . In b) I_e has been dilated and in c) A_{small} has been found. d) shows A_{final} as the dilation of A_{small} .

to close gaps and then search for a closed area on the side of the matched line that is opposite to the mobile robot. When this area has been found the area is dilated in order to compensate for the previous dilation of the edge image. The algorithm is illustrated in Figure 6.

B. Homogeneity Test

We use the initial starting area A_{start} as a training sample and evaluate the rest of the region based on the corresponding colour model. This means that the colour model does not gradually adapt to the growing region, but instead requires a homogeneous region on the complete roof part that is under investigation. Regions that gradually change colour or intensity, such as curved roofs, might then be rejected. However, so far, we did not observe this problem in our experiments.

Gaussian Mixture Models, GMM, are popular for colour segmentation. Like Dahlkamp *et al.* [3] we tested both GMM and a model described by the mean and the covariance matrix in RGB colour space. We selected the mean/covariance model since it is faster and we noted that the mean/covariance model performs approximately equally well as the GMM in our case. A limit O_{lim} is calculated for each model so that 90% of the training sample pixels have a Mahalanobis distance smaller than O_{lim} . O_{lim} is then used as the separator limit between pixels belonging to the class and the pixels that don't.

The result from the local segmentation are regions connected to the lines in L_g , an example is shown in Section VI, Figure 10.

V. GLOBAL SEGMENTATION OF AERIAL IMAGES

In this step the view of the mobile robot is increased further. The previously found building estimates are used as training areas for colour segmentation in order to make a global search for buildings within the entire aerial image. In addition, another important class is introduced, namely driveable areas. The purpose of the global segmentation is to build a map that predicts different types of areas, e.g., driveable ground and buildings. We call this the predictive map, PM. The PM can serve as an input to an exploration algorithm, since it includes both driveable ground and obstacles in the form of buildings.

The global segmentation of an aerial image using colour models captures all buildings with roofs in similar colours as those buildings that were detected in the local segmentation. However, some colours are very similar to ground covered by, e.g., asphalt and ground in deep shadow. Since correct classification is only possible for colours unique to a certain



Fig. 7. The combined binary image of free points and edges in I_e .

class, it is likely that some of the detected building areas may belong to the *ground* class. In order to reduce these false areas, the information about the ground that has been covered by the robot is also used.

A. Colour Models

The segmentation of the aerial image is based on colour models. In the example, models will be calculated for the two classes: *building* and *driveable ground*. We use the same procedure as for the homogeneity test in the local segmentation, see Section IV.

Models of “driveable” ground can be extracted in different ways. Vision has been used in several projects [8], [17], [13] to find driveable regions for unmanned vehicles. In our implementation we use the already available occupancy grid map and interpret free areas as ground. To extract colour models that represent the different ground areas we combine the occupancy grid map and the edge version of the aerial image, I_e . The free cells in the occupancy grid map defines the region in I_e that represents the ground. An example using our occupancy map is shown in Figure 7.

We perform edge controlled segmentation of that region, as described in Section IV-A, to find the different ground areas. The largest areas⁴ point out samples in the aerial image that are used to train colour models of the same type as in Section IV-B.

B. The Predictive Map

The PM is designed to handle multiclass problems and the updating of this map can be performed incrementally. The colour models used to segment the aerial image are two-class models (building or non-building, ground or non-ground etc.) and the classifiers are therefore binary classifiers. We let the PM be a grid map of the same size as the aerial image that is segmented. For each of the n classes, a separate layer l_i , with $i \in \{1, \dots, n\}$, is used to store the accumulated

⁴The limit was set to 50 pixels (12.5 m²) in order to avoid movable objects such as cars and small trucks.

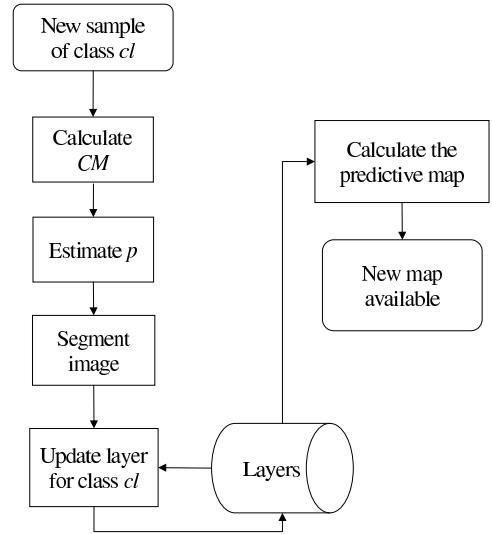


Fig. 8. Flow chart of the process for calculating the predictive map.

segmentation results. These layers also have the same size as the aerial image.

To calculate the predictive map incrementally two main steps are performed; 1) the aerial image is segmented when a new colour model is available and 2) the predictive map is recalculated using the result from the latest segmentation. Figure 8 shows a flow chart of the updating process. This is adapted to work also in an on-line situation and is explained in the following. When a *New sample* belonging to class cl is available a new colour model CM is calculated. Based on the quality of CM , a measure p , $0 \leq p \leq 1$ should be estimated⁵. Then the aerial image is segmented using the new model and the result is scaled with p and stored in a temporary layer. The old layer, l_{cl} , is fused with the temporary layer using a max function.

The predictive map is based on voting from separate layers l_i for the n classes, one layer for each class that we are looking for. In this example $n = 2$; one building layer and one driveable ground layer. The voting is a comparison of the layer’s cell by cell. In the grid cells where the levels are similar, the cells are set to *unknown*. If the unknown areas are classified by the mobile robot as in Figure 9 and 10, that result has precedence, see discussion in Section VI.

The allowed similarity of the cells are defined in C , a matrix where the off-diagonal elements, $c_{ij} \geq 0$, $i \neq j$, $i = \{1, 2, \dots, n\}$, $j = \{1, 2, \dots, n\}$, are used for the classification of cells pm^{xy} in PM. n is the number of layers and classes, and pm^{xy} denotes cell (x, y) in PM. The elements of C introduce buffer zones to the voting process making it possible to adjust the sensitivity of the voting individually for all classes. The voting is performed using IF-THEN rules biased with c_{ij} :

⁵Estimation of the parameter p is still an unsolved issue for future work. In our experiments we used $p = 0.7$.

$$\text{IF } l_i^{xy} > l_j^{xy} + c_{ij} \forall j \neq i \text{ THEN } pm^{xy} = class_i \quad (4)$$

where l_i^{xy} denotes cell (x, y) in layer i . If the condition can not be fulfilled due to conflicting information pm^{xy} is set to *unknown*. If $c_{ij} = 0$ the rules in Equation 4 will turn into ordinary voting where the largest value wins and where ties gives *unknown*.

During the experiments C was set to:

$$C = \begin{bmatrix} - & 0.1 \\ 0.1 & - \end{bmatrix}. \quad (5)$$

VI. EXPERIMENTS

A. Data Collection

The above presented algorithms have been implemented in Matlab for evaluation and the functions currently work off-line. Data were collected with a mobile robot, a Pioneer P3-AT from ActivMedia, equipped with differential GPS, laser range scanner, cameras and odometry. The robot is equipped with two different types of cameras; an ordinary camera mounted on a PT-head and an omni-directional camera. The omni-directional camera gives a 360° view of the surroundings in one single shot. The camera itself is a standard consumer-grade SLR digital camera (Canon EOS350D, 8 megapixels). On top of the lens, a curved mirror from 0-360.com is mounted. From each omni-image we compute 8 (every 45°) planar views or sub-images with a horizontal field-of-view of 56°. These sub-images are the input to the virtual sensor. The images were taken with ca. 1.5 m interval and were stored together with the corresponding robot's pose. The trajectory of the mobile robot is shown in Figure 3.

B. Tests of the Local Segmentation

The occupancy map in Figure 9 was built using the horizontally mounted laser range scanner. The occupied cells in this map (marked in black) were labelled by the virtual sensor giving the semantic map presented in Figure 1. The semantic map contains two classes; buildings (values above 0.5) and non-buildings (values below 0.5). From this semantic map we extracted the grid cells with a high probability of being a building (above 0.9) and converted them to the lines L_g^M presented in Figure 2. Matching of these lines with the lines extracted from the aerial image L_a^N , see Figure 4, was then performed. Finally, based on best line matches the segmentation was performed according to the description in Section IV.

The three parameters in \mathbf{R} (Equation 3) were set to $\sigma_{Rx} = 1$ m, $\sigma_{Ry} = 1$ m, and $\sigma_{R\theta} = 0.2$ rad. Note that it is only the relation between the parameters that influences the line matching.

We have performed two different types of tests. *Tests 1-3* are the nominal cases when the collected data are used as they are. The tests intend to show the influence of a changed relation between σ_{Rx} , σ_{Ry} and $\sigma_{R\theta}$ by varying $\sigma_{R\theta}$. In *Test 2* $\sigma_{R\theta}$ is decreased by a factor of 2 and in *Test 3* $\sigma_{R\theta}$ is increased by a factor of 2. In *Tests 4* and *5* additional



Fig. 9. Occupancy map used to build the semantic map in Fig. 1.

uncertainty (in addition to the uncertainty already present in L_g^M and L_a^N) was introduced. This uncertainty is in the form of Gaussian noise added to the midpoints (σ_x and σ_y) and directions σ_θ of L_g^M . The tests are defined in Table I.

Test	σ_x [m]	σ_y [m]	σ_θ [rad]	$\sigma_{R\theta}$ [rad]	N_{run}
1	0	0	0	0.2	1
2	0	0	0	0.1	1
3	0	0	0	0.4	1
4	1	1	0.1	0.2	20
5	2	2	0.2	0.2	20

TABLE I
DEFINITION OF TESTS AND THE USED PARAMETERS.

C. Quality Measure

We introduce two quality measures to be able to compare different algorithms or sets of parameters in an objective way. For this four sets (A-D) are defined; A is the ground truth, a set of cells/points that has been manually classified as the tested *class*; B is the set of cells that has been classified as *class* by the algorithm; C is the set of false positives, $C = B \setminus A$, the cells that have been classified as *class B* but do not belong to ground truth *A*; and D are the true positives, $D = B \cap A$, the cells that have been classified as *class B* and belong to ground truth *A*. Using these sets, two quality measures are calculated as:

- The true positive rate, $\Phi_{TP} = \#D/\#B$.
- The false positive rate, $\Phi_{FP} = \#C/\#B$.

where $\#D$ denotes the number of cells in *D*, etc.

D. Result of Local Segmentation

The results of *Test 1* show a high detection rate (96.5%) and a low false positive rate (3.5%), see Table II. The resulting segmentation is presented in Figure 10. Four deviations from an ideal result can be noted. At *a* and *b* tree tops are obstructing the wall edges in the aerial image, at *c* a white wall causes a gap between two regions, and a false area, to the left of *b*, originates from an error in the semantic map (a low hedge was marked as building).

The results of *Test 1-3* are very similar which indicate that the algorithm in this case was not specifically sensitive

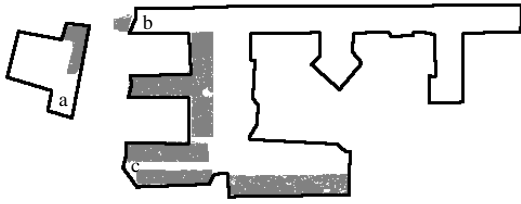


Fig. 10. The result of the local segmentation of the aerial image using the wall estimates in Figure 2. The ground truth building outlines are drawn in black.

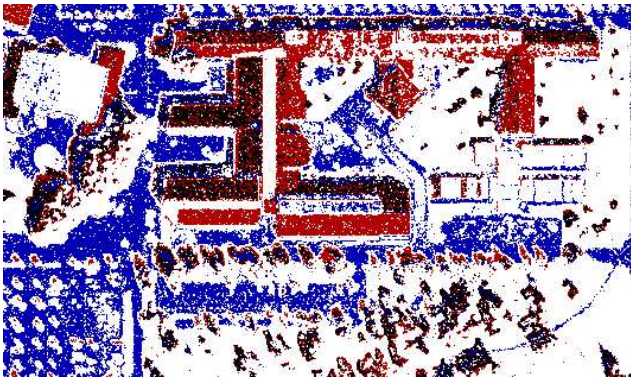


Fig. 11. The result of the global segmentation of the aerial image using both ground and building models. Ground is blue, buildings are red, ties are black and white represents not classified cells.

to the changes in $\sigma_{R\theta}$. In *Test 4* and *5* the scenario of *Test 1* was repeated using a Monte Carlo simulation with introduced pose uncertainty. The result is presented in Table II. One can note that the difference between the nominal case and *Test 4* is very small. In *Test 5* where the additional uncertainties are higher the detection rate has decreased slightly.

Test	Φ_{TP} [%]	Φ_{FP} [%]
1	96.5	3.5
2	97.0	3.0
3	96.5	3.5
4	96.8 ± 0.2	3.2 ± 0.2
5	95.9 ± 1.7	4.1 ± 1.7

TABLE II

RESULTS FOR THE TESTS. THE RESULTS OF TEST 4 AND 5 ARE PRESENTED WITH THE CORRESPONDING STANDARD DEVIATION.

E. Result of Global Segmentation

The result of the global segmentation has been assessed mainly by visual inspection. The visual inspection of the result shown in Figures 11 and 12, and of segmentation of a larger aerial image illustrates the potential in the use of aerial images for mapping purposes. The PM based on ground colour models from regions in Figure 7 and building colour models from the regions in Figure 10 is presented in Figure 11.

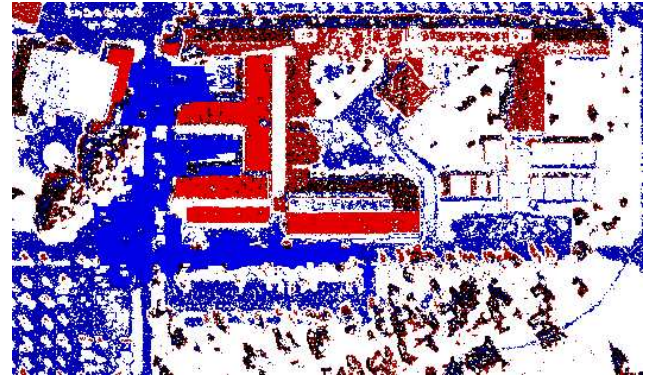


Fig. 12. PM combined with the local information. Ground is blue, buildings are red, ties are black and white represents not classified cells.

Compared with the aerial image in Figure 3 the result is promising. One can now follow the outline of the main building and most of the paths, both paved paths and roads and beaten tracks have been found. The major problem that has been noted during the work is caused by shadowed areas that look very similar to dark roofs.

The final result is obtained when the PM is combined with the free areas and the buildings found in the local segmentation. For these pixels we set $p = 0.9$, performed another update (where segmentation is not needed) and got the resulting map shown in Figure 12.

A formal evaluation of the ground class is hard to perform. Ground truth for buildings can be manually extracted from the aerial image, but it is hard to specify in detail the areas that are driveable. Based on the ground truth of buildings and an approximation of the ground (driveable areas) ground truth as the non-building cells, some statistics of the result are presented in Table III. In the table all values in the right column, where the results from the combined PM and local information are shown, are better than those in the middle column. The increase in true positive rate for buildings is depending both on the reduction of *ties* and on the reduction of false positives on the road along the mobile robot trajectory. Since the result depends on the relation between the actual presence of the different classes in the aerial image, normalized values for Φ_{TP} are also presented. The area covered by buildings is smaller than the ground area giving an increase in the normalized Φ_{TP} for buildings and a decrease for ground compared to the nominal Φ_{TP} .

Description	PM [%]	PM + local [%]
Φ_{TP} buildings (normalized)	66.6 (83.2)	73.0 (87.7)
Φ_{TP} ground (normalized)	96.8 (92.4)	97.3 (93.3)
Unclassified cells	55.5	52.4
Ties (unknown)	10.5	8.1

TABLE III

RESULTS OF THE EVALUATION OF THE TWO PMs DISPLAYED IN FIG. 11 AND 12. NOTE THAT THE GROUND TRUTH FOR 'GROUND' IS AN APPROXIMATION AND THESE VALUES WOULD DECREASE USING THE EXACT GROUND TRUTH.

VII. CONCLUSIONS AND FUTURE WORK

This paper discusses how aerial images can be used to considerably extend the view of a mobile robot. A virtual sensor for building detection on a mobile robot is used to link semantic information to a process for building detection in aerial images. The benefit from the extended range of the robot's view can clearly be noted in the presented example.

In the local segmentation it can be hard to extract a complete building outline due to, e.g., different roof materials, different roof inclinations and additions on the roof, specifically when the robot has only seen a small portion of the building outline. But the global segmentation is a promising extension that shows a large potential. Even though the roof structure in the example is quite complicated, the outline of a large building could be extracted based on the limited view of the mobile robot, which had only seen a minor part of surrounding walls.

A. Discussion

With the presented method, changes in the environment compared to an aerial image that is not perfectly up-to-date are handled automatically. Assume that a building that is present in the aerial image has been removed after the image was taken. Since it is present in the aerial image, it may be classified as a building in the PM if it had a roof colour similar to a building already detected by the mobile robot. When the robot approaches the area where the building was situated, the building will not be detected. If the mobile robot classifies the area as driveable ground, the PM will turn into *unknown* (of course depending on c_{ij} and p), not only for that specific area but also globally. Still, the information obtained from the local segmentation will indicate building where the mobile robot initially found the building that resulted in the first colour model.

What about the other way around? Assume that a new building is set up on an open area and this is not yet reflected in the aerial image. If the edge matching indicates a wall this can introduce errors, but if the building is in the middle of, e.g., a lawn, no edges are found and no segmentation will be performed. Then the building will only be present in the probabilistic semantic map in the form of a possible wall.

B. Future Work

We believe that the accuracy of the PM could be further improved by using a measure of the colour model quality to assign a value to the parameter p for each model. Also the probabilities from the semantic map where the ground wall estimates are extracted could be included in the calculation of p .

Shadow detection that merges areas in shadows with the corresponding areas in the sun is desired. We believe that this would eliminate some false pixels and decrease the unknown areas caused by ties.

Experiments where PM is used to direct exploration of unknown areas should be performed. At the same time it should be investigated whether post-processing of the PM can increase the detection rates further.

REFERENCES

- [1] J. Canny. A computational approach for edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(2):279–98, Nov 1986.
- [2] C. Chen and H. Wang. Large-scale loop-closing with pictorial matching. In *Proceedings of the 2006 IEEE International Conference on Robotics and Automation*, pages 1194–1199, Orlando, Florida, May 2006.
- [3] H. Dahlkamp, A. Kaehler, D. Stavens, S. Thrun, and G. Bradski. Self-supervised monocular road detection in desert terrain. In *Proceedings of Robotics: Science and Systems*, Cambridge, USA, June 2006.
- [4] J. Freixenet, X. Munoz, D. Raba, J. Marti, and X. Cufi. Yet another survey on image segmentation: Region and boundary information integration. In *European Conference on Computer Vision*, volume III, pages 408–422, Copenhagen, Denmark, May 2002.
- [5] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.
- [6] R. C. Gonzales and R. E. Woods. *Digital Image Processing*. Prentice-Hall, 2002.
- [7] J. Guerrero and C. Sagüés. Robust line matching and estimate of homographies simultaneously. In *Pattern Recognition and Image Analysis: First Iberian Conference, IbPRIA 2003*, pages 297–307, Puerto de Andratx, Mallorca, Spain, June 2003.
- [8] Y. Guo, V. Gerasimov, and G. Poulton. Vision-based drivable surface detection in autonomous ground vehicles. In *Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3273–3278, Beijing, China, Oct 9-15 2006.
- [9] H. Mayer. Automatic object extraction from aerial imagery – a survey focusing on buildings. *Computer vision and image understanding*, 74(2):138–149, May 1999.
- [10] M. Mueller, K. Segl, and H. Kaufmann. Edge- and region-based segmentation technique for the extraction of large, man-made objects in high-resolution satellite imagery. *Pattern Recognition*, 37:1621–1628, 2004.
- [11] S. M. Oh, S. Tariq, B. N. Walker, and F. Dellaert. Map-based priors for localization. In *IEEE/RSJ 2004 International Conference on Intelligent Robotics and Systems*, pages 2179–2184, Sendai, Japan, 2004.
- [12] M. Persson, T. Duckett, and A. Lilienthal. Virtual sensor for building detection by an outdoor mobile robot. In *Proceedings of the IROS 2006 workshop: From Sensors to Human Spatial Concepts*, pages 21–26, Beijing, China, Oct 2006.
- [13] M. Persson, T. Duckett, and A. Lilienthal. Improved mapping by using semantic information to fuse overhead and ground-level vision. In *The 13th International Conference on Advanced Robotics, ICAR*, Jeju, Korea, Aug 21-14 2007.
- [14] M. Persson, T. Duckett, C. Valgren, and A. Lilienthal. Probabilistic semantic mapping with a virtual sensor for building/nature detection. In *The 7th IEEE International Symposium on Computational Intelligence in Robotics and Automation CIRA2007*, June 21-24 2007.
- [15] C. Scrapper, A. Takeuchi, T. Chang, T. H. Hong, and M. Shneier. Using a priori data for prediction and object recognition in an autonomous mobile vehicle. In G. R. Gerhart, C. M. Shoemaker, and D. W. Gage, editors, *Unmanned Ground Vehicle Technology V. Edited by Gerhart, Grant R.; Shoemaker, Charles M.; Gage, Douglas W. Proceedings of the SPIE, Volume 5083, pp. 414-418 (2003).*, pages 414–418, Sept. 2003.
- [16] D. Silver, B. Sofman, N. Vandapel, J. A. Bagnell, and A. Stentz. Experimental analysis of overhead data processing to support long range navigation. In *Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2443–2450, Beijing, China, Oct 9-15 2006.
- [17] D. Song, H. N. Lee, J. Yi, and A. Levandowski. Vision-based motion planning for an autonomous motorcycle on ill-structured road. In *Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3279–3286, Beijing, China, Oct 9-15 2006.
- [18] F. Tupin and M. Roux. Detection of building outlines based on the fusion of SAR and optical features. *ISPRS Journal of Photogrammetry & Remote Sensing*, 58:71–82, 2003.
- [19] I. Ulrich and I. Nourbakhsh. Appearance-based obstacle detection with monocular color vision. In *AAAI National Conference on Artificial Intelligence*, pages 866–871, Austin, TX, July 30 - Aug 3 2000.