ROBUST EXTRACTION OF TRAFFIC SIGNS FROM GEOREFERENCED MOBILE MAPPING IMAGES

Alexander Reiterer*, Taher Hassan*, Naser El-Sheimy*

* Institute of Geodesy and Geophysics, Research Group Engineering Geodesy, Vienna University of Technology, Gusshausstr. 27-29 / E128-3, A-1040 Vienna. alexander.reiterer@tuwien.ac.at
http://info.tuwien.ac.at/ingeo

b Mobile Multi-Sensor Research Group, Department of Geomatics Engineering, The University of Calgary, 2500 University Drive N.W., Calgary, Alberta, Canada. ttabas@ucalgary.ca, elsheimy@ucalgary.ca
http://www.geomatics.ucalgary.ca

KEY WORDS: Driver Assistance System, Mobile Mapping System, Scale Invariant Feature Transform.

ABSTRACT:

Modern Driver Assistance Systems (DAS) are required to assist, guide, and control vehicles on highways and city streets based on GNSS, INS and map matching. They play an important role in modern vehicles navigation. Although a GNSS-navigation system can be updated in view of the modifications of the roads, it does not include exhaustive information about the traffic signalization. It would be useful to signal to a driver at least some important traffic signs. This paper presents the basic concept of a new approach for the automated detection of traffic signs to be incorporated in DASs. The developed procedure is based on the well known Scale Invariant Feature Transform (SIFT) algorithm. The results of extensive testing on real data sets show that the presented approach detects and classifies over 70% of traffic signs correctly.

1 INTRODUCTION

Driver Assistance Systems (DAS) play an important role in modern vehicles navigation(Kumar, 1997). Such systems are required to assist drivers and provide guidance and probably control of their vehicles on highways and city streets. Today, drivers are already helped by automatic systems, e.g. GNSS/digital maps navigation systems. However, although a GNSS-navigation system can be updated in view of the modifications of the roads, it does not include exhaustive information about the traffic signalization. The main reason is that traffic signalization may change without notice.

To effectively assist drivers, it would be useful to signal at least some important traffic signs. This will potentially avoid accidents due to driver’s negligence or miss-consciousness. The main problem with including traffic signs into such an assistance system is that collecting traffic signs location/information by conventional methods is very expensive and time consuming. These methods are therefore not well suited for rapid updating of existing traffic sign databases.

2 MOBILE MAPPING SYSTEMS

Land-based Mobile Mapping Systems (MMS) have yielded an enormous time saving in road networks and their surrounding utilities surveys (Schwarz and El-Sheimy, 2004). However, the manual extraction of the road information from the mobile mapping images is still a time-consuming task. The main goal of an international collaboration between the University of Calgary (Department of Geomatics Engineering) and the Vienna University of Technology (Institute of Geodesy and Geophysics) is the development of a new technique for automatic traffic signs recognition from geo-referenced mobile mapping image sequences. Comparable developments can be found in (Escalera and Salichs, 1997, Kumar, 1997, Lalonde, 1995, Prince, 1998, Zadeh and Suen, 1997, Zhiyuan, 1998). The advantage of the developed technique is, that (1) the whole procedure is fully integrated into a running MMS system, (2) the procedure is real-time capable and (3) the procedure is easily extensible.

The developed technique has been tested on the VISAT™ mobile mapping system image sequences (Figure 1). VISAT™ is developed by the department of Geomatics Engineering (University of Calgary) in cooperation with Absolute Mapping Solution (AMS).

Figure 1: VISAT™ Van (Absolute Mapping Solutions).

The hardware components of the VISAT™ mobile mapping system are a Strap Down Inertial Navigation System (SINS), a dual-frequency GNSS receiver, and a cluster of digital color cameras. The primary purposes of these components are:
• provide the position/orientation by GNSS/SINS integration,
• using the camera cluster for relative positioning.

However, these components also have important secondary functions. The GNSS controls the long-term error growth of the SINS through the GPS/SINS Kalman filter and provides the precise timing base for all data streams. The SINS is used as a position sensor in addition to an orientation sensor – consequently, these tasks include bridging GPS signal outages, detecting and correcting GNSS cycle slips, and precise interpolation between GNSS positions. The latter task – interpolation between GNSS positions – is possible because the SINS provide data at 200 Hz, while the GNSS positions and velocities are only available at 1-10 Hz. Actual position and time are used for tagging the captured images – the result is a geo-referenced image database.

In addition to the GNSS, SINS, and cameras, the VISAT™ system also integrates a Distance Measuring Instrument (DMI). The DMI is used to trigger the acquisition of the images from the cameras at constant distance intervals defined by the user. Figure 2 shows the VISAT™ hardware.

![Figure 2: VISAT™ system components.](image)

As mentioned above the system includes a camera cluster which consists of two camera enclosures. Each of these enclosures is attached on a side of the roofmount (each camera enclosure is the exact mirror of the other). The field of view of the whole camera cluster is 330°. The developed system is based on images captured by cameras 1 and 5 which face forward and contain most of the traffic signs (see Figure 3). In a later step the use of the other cameras is envisaged.

![Figure 3: Camera’s field of view.](image)

3 THE DEVELOPED DETECTION PROCEDURE

For the development and the prospective productive use of the procedure, several preconditions and design criteria can be formulated:

• fast execution (nearly real time),
• easy to use and easy to include new data sets (new traffic sign models/templates).

Due to these preconditions, the problem has been reduced to a matching problem between pre-selected traffic signs, forming a template database, and the traffic signs extracted from geo-referenced mobile mapping image sequences. Nevertheless, the matching between a huge database of traffic signs and new images is an intensive task.

The whole procedure for the automatic detection of traffic signs can be divided into the following steps (see Figure 4):

• building the standard traffic signs database (consisting of traffic sign models/templates),
• extracting image feature descriptors for database models,
• extracting MMS images feature descriptors,
• matching descriptors of the images with all descriptors in the database,
• clustering the resulting matches to get the separated traffic signs (if present), and
• calculating the 3D coordinates of traffic signs using forward intersection (if necessary).

Building the traffic sign models and extracting the corresponding image feature descriptors can be done in a pre-processing step (independent from the on-line procedure).

![Figure 4: Developed detection procedure.](image)

As indicated above, the last two steps are optional. Clustering of traffic signs has to be done if more than one traffic sign is present in the scene, calculation of 3D coordinates represents an additional information for mapping (if desired).

3.1 Building the Database of Typical Traffic Signs

As mentioned before, the whole procedure is based on the matching between traffic signs which have been selected in an independent working step and images which are captured by the DAS camera(s). Therefore, as a first step the database has to be filled by suitable traffic sign models. These models are cropped images which are extracted out from real VISAT™ MMS images.
Each type of traffic sign has to be represented in the database at least by one pre-selected traffic sign model. Figure 5 shows some examples of traffic sign models.

Cropping traffic sign models can be done by a conventional image processing software (e.g. Gimp, Photoshop, or by batch command line image crop). Tests have shown that best results can be achieved using traffic sign models of high image quality (constant illumination, low image noise, etc.).

### 3.2 Calculating Image Features of Models and of the Actual Image Pair

For the description of images by features, a huge number of different algorithms exists, e.g. histogram features (Pratt, 1978), Haralick moments (Haralick and Shapiro, 1993), etc.

For the presented approach we have used the Scale Invariant Feature Transform (SIFT) algorithms developed by Lowe (Lowe, 2004). This algorithm has some characteristics, which are significant for the automatic extraction task, e.g. extracted features are reasonably invariant to changes in illumination, image noise, rotation, scaling, and (small) changes in viewpoint. The SIFT algorithm works on the basis of several steps: detection of extrema in scale-space, localization of keypoints, assignment of orientation, and the generation of descriptors. The detection of extrema in the scale space is done by means of the calculation of Difference-of-Gaussian (DoG) at different scales. Keypoints are identified as local maxima or minima of the DoG images across different scales. Each pixel in the DoG images is compared to its 8 neighbours at the same scale, plus the 9 corresponding neighbours at neighbouring scales. If a pixel is a local maximum or minimum, it is selected as a candidate keypoint. To determine the orientation of the keypoints in their neighbourhood, gradient orientation histograms are computed. The contribution of each neighbouring pixel is weighted by the gradient magnitude and a Gaussian window with a size that is 1:5 times the scale of the keypoint. Peaks in the histogram correspond to dominant orientations. On the basis of the calculated keypoints and their orientation (see Figure 6), the feature descriptors can be determined.

![Figure 6: Sift descriptor representation.](image)

The Sift features are a set of orientation histograms on $4 \times 4$ pixel neighbourhoods. A histogram consists of 8 bins each, and each descriptor contains an array of 4 histograms around the keypoint. This leads to a SIFT feature vector with $4 \times 4 \times 8 = 128$ elements. This vector is normalized to enhance invariance to changes in illumination. Such a vector is calculated for each extrema in scale-space – the result can be seen as a set of interest points and their feature vectors. The calculation of the features of the models has to be done uniquely – features are stored with all model data in a database.

### 3.3 Matching the Descriptors of the Images with all Descriptors in the Database

To find traffic signs in the MMS images, the calculated image features have to be compared with the features in the database. For this process, a suitable matching routine has to be used. We are investigating the matching process as proposed by Lowe (Lowe, 2004). This process is based on finding a match between image features and model features by evaluating the Euclidean distance. According to the Nearest Neighborhood procedure for each $F_i$ feature in the model image feature set the corresponding feature $F_j^2$ must be looked for in the model feature database. The corresponding feature is one with the smallest Euclidean distance to the feature $F_i^2$. A pair of corresponding features is called a match $(F_i^1, F_j^2)$. If the Euclidean distance between the two features $F_i^1$ and $F_j^2$ is below a certain threshold, the match $M(F_i^1, F_j^2)$ is labelled as positive.

An important issue for the matching routine is how the reliability and runtime varies as a function of the number of features. Due to our system is in prototype status no fundamental research has been done for evaluating runtime. More information about this issue can be found under (Lowe, 2004).

The developed procedure improves the robustness of the matching process by evaluating the distance of at least two matches. As a result, a list of points, which have been classified as points inside a traffic sign, are obtained.

### 3.4 Clustering the Resulting Matches to Get Separated Traffic Signs

The matches between the database and the actual feature points, which have been classified as points inside a traffic sign, have to be clustered. This part is necessary to separate different point groups and to collect them to single traffic sign point clusters. For this part, the k-means clustering algorithm is used. This algorithm is a very effective method to group objects based on attributes. The grouping is processed by minimizing the sum of squares of distances between data and the corresponding cluster centroid. The k-means algorithm for our problem needs two input parameters: the expected number of clusters (this is a characteristic of the k-means clustering) and the minimal distance between separated clusters (traffic signs). The expected number of clusters can be determined by processing the k-means procedure twice. In the first run the number of clusters is fixed to a number which is certainly higher than the expected maximum number of traffic signs in one image (this parameter is set depending on the available image sets – we have fixed the parameter to 5). A result of this first run is the number of present clusters – this number can be used as input for the second run of the k-means algorithm (as an alternative a combination of hierarchical clustering and k-means is often proposed in literature (Hartigan and Wong, 1979)). The main advantages of the k-means procedure are simplicity and speed. The result of the clustering is the above mentioned point list which is now extended by a cluster number for each point. This step is processed for both images (for camera 1 and camera 5). In ideal case, the number of traffic signs (clusters) is equal in both images. On the basis of the image coordinates of the corresponding traffic signs (cluster centers) the calculation of the 3D coordinates by forward intersection is possible.
4 EVALUATION

The developed procedure has been tested on about 800 images (400 image pairs), which have been the result of two different survey-drives. One sequence was collected under very difficult illumination conditions – the resulting images can be described as under-exposed and therefore as generally too dark (see Figure 7). The second image sequence was taken under ideal light conditions – images are well illuminated (see Figure 8).

Figure 7: Example for an image from the sequences which are used for evaluating the developed procedure. The yellow rectangle shows the detected traffic sign.

Figure 8: Example for an image from the sequences which are used for evaluating the developed procedure. The yellow rectangle shows the detected traffic sign.

The database of models has been built in pre-processing consisting of independent images. This means that the models are cropped out from images which are taken by the MMS camera system – the used images are not part of the testing sequence.

<table>
<thead>
<tr>
<th>Traffic Signs</th>
<th>Total</th>
<th>Camera 1</th>
<th>Camera 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detected Traffic Signs</td>
<td>120</td>
<td>64</td>
<td>56</td>
</tr>
<tr>
<td>Detected Traffic Signs [%]</td>
<td>78%</td>
<td>81%</td>
<td>73%</td>
</tr>
</tbody>
</table>

Table 1: Results of image sequence 1 (poor quality).

<table>
<thead>
<tr>
<th>Traffic Signs</th>
<th>Total</th>
<th>Camera 1</th>
<th>Camera 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detected Traffic Signs</td>
<td>134</td>
<td>66</td>
<td>68</td>
</tr>
<tr>
<td>Detected Traffic Signs [%]</td>
<td>72%</td>
<td>69%</td>
<td>76%</td>
</tr>
</tbody>
</table>

Table 1 shows that in total 72% of the traffic signs have been detected. Evaluating the detection rate separated for each camera results in 69% for the left camera (camera 1) and 76% for the right camera (camera 5). The significant difference in the detection rate between the two cameras can be explained by a high difference of the sharpness of the two camera’s images (see Figure 9).

The difference in the sharpness is presented in all image sequences, but seems to be stronger for the images captured under bad/poor lightening conditions (e.g. low light). Image sequence 2 (see Table 2) shows a significant higher detection rate (left camera / camera 1: 81%; right camera / camera 5: 73%). In this image sequence a considerable difference of the detection rate between the two cameras is also present (reverse to image sequence 1) – an explanation for this effect has not been found until yet. Table 3 shows the detection rate referred to the entire benchmark test (all two image sequences together) – all detection rates are around 74%.

Table 2: Results of image sequence 2 (ideal quality).

<table>
<thead>
<tr>
<th>Traffic Signs</th>
<th>Total</th>
<th>Camera 1</th>
<th>Camera 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detected Traffic Signs</td>
<td>227</td>
<td>118</td>
<td>109</td>
</tr>
<tr>
<td>Detected Traffic Signs [%]</td>
<td>74%</td>
<td>74%</td>
<td>74%</td>
</tr>
</tbody>
</table>

Table 3: Results (all image sequences together).

For the evaluation we have defined that only traffic signs which exceed the dimension of $20 \times 20$ pixels in both images have to be detected. Traffic signs which have a smaller dimension are neglected and will be detected from a closer pair. Table 1, 2, 3 show the evaluation results.

Generally, the detection system works in a more than satisfactory way – more than 70% of the traffic signs can correctly be detected and classified. The result highly depends on the models, included into the system database and on the quality of the image sequences. Due to system design flexibility, the inclusion of more models becomes not an issue – a user can crop the concerning traffic sign and copy them into the appropriate directory.

To have a sufficient image quality the capturing process has to be done under good light conditions. Poor image quality can be improved by means of pre-processing (histogram equalization and median filtering should give adequate results).
5 CONCLUSION

The article presents an automated traffic sign detection method which has been tested on image sequences captured by a mobile mapping system (VISAT\textsuperscript{TM}). The presented technique is based on existing algorithms, like the SIFT operator and the k-means clustering. The developed procedure uses a conventional matching routine combined with a database to detect traffic signs in MMS image sequences. This approach has been chosen due to the possibility to include new traffic sign models and the possibility to realize the whole procedure as real-time process.

The whole sequence has been implemented under MATLAB into a running prototype. Tests have been processed by a huge number of examples of different quality (over 800 images) – the system shows a sufficient detection rate of about 74%. Under ideal lighting conditions (e.g. constant illumination) detection rates of over 80% have been achieved.

Future work will extend the algorithm to a real-time working procedure. To realize this task we plan a re-implementation of the whole process in C\#. The detection/ extraction of image features (SIFT) can be implemented using the Graphical Processing Unit (GPU).

Furthermore, extensive tests will be necessary for a detailed evaluation of the developed procedure.

ACKNOWLEDGEMENTS

The presented work is supported by the Vienna University of Technology (Austria) and the University of Calgary (Canada).

REFERENCES


