Automated Traffic Sign Detection for Modern Driver Assistance Systems

Alexander REITERER, Austria
Taher HASSAN, Naser EL-SHEIMY, Canada

Key words: Driver Assistance System, Mobile Mapping System, Scale Invariant Feature Transform.

SUMMARY

Modern Driver Assistance Systems (DAS) are required to assist, guide, and control vehicles on highways and city streets based on GPS, INS and map matching. They play an important role in the navigation of modern vehicles. Although a GPS-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 shows that the presented approach detects 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. A more advanced technique represents the on-line detection and classification of traffic signs.

In recent years a huge number of papers concerning traffic sign detection have been published. Piccioli et al. (1996) use color and priori information to limit the possible locations of signs in the image. They are extracting edges and are searching for circular or triangular regions before applying a cross-correlation technique for recognizing the signs. In Estevez and Kehtarnavaz (1996) a measure is used to locate stop, yield, and “do-not-enter” signs. This step is followed by edge detection and shape analysis to identify the sign. Escalera et al. (1997) start with color matching and corner detection for analyzing corners in specific relationships that correspond to triangular, rectangular, or circular signs. Classification makes use of a neural network. Yuille et al. (1998) are using correction of the colors of the ambient illumination and are mapping the detected sign into a front parallel position before analyzing them. Huang and Hsu (2000) use shape and color in a wide angle view to locate signs as circular or triangular shapes. Paclik et al. (2000) is using a decision tree method to detect and recognize signs without using color. Detection is based on shape using local orientations of image edges and hierarchical templates. The results are analyzed by a decision tree which either labels the regions by sign type or rejects them. In Sekanina et al. (2002) speed limit signs are detected using color and multi-resolution techniques. Fleischer et al. (2002) shows a very different approach – the developed method uses a model-based, top down process. Predictions are made for locations in which signs may appear. Shape, size, and color are used to specify edge models for signs in the 3D world. Signs that are found are tracked through subsequent images using a Kalman filter. Shaposhnikov et al. (2002) make use of color segmentation using the CIECAM97 color appearance model. Histograms of oriented edge
elements are used to determine the shape of the sign followed by location of the center of the sign. Signs are described by a set of descriptors which are matched with stored models to recognize the signs. Fang et al. (2003) is focusing on sign detection in cluttered environments. Neural networks are used to locate regions that may be the center of signs (using color and shape). Candidate signs are tracked through time using a Kalman filter. Other methods and developments for the detection of traffic signs have been published by Kumar (1997), Lalonde (1995), Prince (1998), Zadeh et al. (1997), Hsien et al. (2003), Shadeed et al. (2003), Gao et al. (2003), Oh et al. (2005), Marmo et al. (2006).

The presented approach follows the general trend in previous work. The identification and recognition of traffic signs is realized using feature descriptors and a robust matching routine. Traffic signs can be tracked over time analyzing image sequences. The details of the approach are different, and the method has the advantages of being fast and easily modified to recognize new classes of signs. Furthermore, the method has been implemented and tested for a Land-based Mobile Mapping Systems.

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.

The developed technique has been tested on the VISAT™ mobile mapping system (Figure 1a gives a general overview). VISAT™ is developed by the department of Geomatics Engineering (University of Calgary) in cooperation with Absolute Mapping Solution (AMS). 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 1b shows the VISAT™ hardware.

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 2). In a later step the use of the other cameras is envisaged.

**Figure 2:** Camera’s field of view.
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 – for our application traffic sign detection for one image pair should be processed in under 1 sec.),
- 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 3):
- 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 3: Developed detection procedure.
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).

As mentioned in the Introduction (Section 1) there is a huge number of techniques and methods for the detection of traffic signs. The development has been focused on using standard image and learning algorithms and on the flexibility of the whole procedure (including new traffic sign models for detection should not be a big deal – the possibility for a fast and easy extension is requested).

3.1 Building the Database of Typical Traffic Signs

The developed 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.

![Figure 4: Examples of traffic sign models.](image)

Each type of traffic sign has to be represented in the database at least by one pre-selected traffic sign model. Figure 4 shows some examples of traffic sign models. In the United States, the Manual on Uniform Traffic Control Devices defines the appearance of all signs and road markings (see: [http://mutcd.fhwa.dot.gov/pdfs/2003r1/pdfindex.htm](http://mutcd.fhwa.dot.gov/pdfs/2003r1/pdfindex.htm)). In a later step the integration of models on the basis of this manual is envisaged.

Cropping traffic sign models can be done by a conventional image processing software (e.g. Gimp, Photoshop, or by batch command line image crop). The presented approach works on the basis of image descriptors for representing the models and for the later matching process. Therefore it is very important to include the traffic sign models in a sufficient quality (constant illumination, low image noise, etc.) and in an adequate size (for our experiments we have included traffic signs of a minimal size of 50 x 50 pixel).

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 (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 key-points, 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. Key-points 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 neighbors at the same scale, plus the 9 corresponding neighbors at neighboring scales. If a pixel is a local maximum or minimum, it is selected as a candidate key-point. To determine the orientation of the key-points in their neighborhood, gradient orientation histograms are computed. The contribution of each neighboring pixel is weighted by the gradient magnitude and a Gaussian window with a size that is 1:5 times the scale of the key-point. Peaks in the histogram correspond to dominant orientations. On the basis of the calculated key-points and their orientation (see Figure 5), the feature descriptors can be determined.

![Figure 5: SIFT descriptor representation.](image)

The SIFT features are a set of orientation histograms on 4×4 pixel neighborhoods. A histogram consists of 8 bins each, and each descriptor contains an array of 4 histograms around the key-point. This leads to a SIFT feature vector with 4×4×8 = 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 (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 $F_2^j$ must be looked for in the model feature...
database. The corresponding feature is one with the smallest Euclidean distance to the feature $F_i^i$. A pair of corresponding features is called a match $(F_i^i, F_j^j)$. If the Euclidean distance between the two features $(F_i^i$ and $F_j^j$) is below a certain threshold, the match $M(F_i^i, F_j^j)$ is labeled 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 done for evaluating runtime. More information about this issue can be found under (Lowe, 2004).

To improve the robustness of the matching process single points are not classified as points inside a traffic sign. The developed procedure analyses the distance of at least two matches. If the distance is less than a predefined value points are classified as points inside a traffic sign. For our data set we have fixed this value to 50 pixels.

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. Lowe (2004) proposed a combination of Hough transformation and RANSC algorithm for finding the correct pose from cluttered matches. For our problem this approach would be oversized. Therefore we use a simple but very powerful and fast clustering method: k-means. Furthermore k-means can be implemented as a real-time application (Maliatski & Yadid-Pecht 2005, Kanungo et al. 2002).

K-means 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 (an example is shown in Figure 6a). The second image sequence was taken under ideal light conditions – images are well illuminated (an example is shown in Figure 6b).

![Figure 6a](image)
![Figure 6b](image)

**Figure 6:** Two examples for images 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’s system – the used images are not part of the testing sequence.

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Camera 1</th>
<th>Camera 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traffic Signs</td>
<td>186</td>
<td>96</td>
<td>90</td>
</tr>
<tr>
<td>Detected Traffic Signs</td>
<td>134 (72%)</td>
<td>66 (69%)</td>
<td>68 (76%)</td>
</tr>
<tr>
<td>Not Detected Traffic Signs</td>
<td>52 (28%)</td>
<td>30 (31%)</td>
<td>22 (24%)</td>
</tr>
<tr>
<td>False Detections</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

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

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.

Table 1 show that under difficult illumination conditions 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 8a and b). For this survey drive three traffic signs have been detected false (detection of a traffic sign where no traffic sign is). This number can be described as very low and emphasizes the potential of the presented approach. The false detected objects are mainly located near dense vegetation (e.g. trees, bushes, etc.).

<table>
<thead>
<tr>
<th>Total</th>
<th>Camera 1</th>
<th>Camera 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traffic Signs</td>
<td>120</td>
<td>64</td>
</tr>
<tr>
<td>Detected Traffic Signs</td>
<td>93 (78%)</td>
<td>52 (81%)</td>
</tr>
<tr>
<td>Not Detected Traffic Signs</td>
<td>27 (22%)</td>
<td>12 (19%)</td>
</tr>
<tr>
<td>False Detections</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

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

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 (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). The false detection rate is nearly in the same magnitude as in the previous survey drive (total number of 2) and therefore still impressive. Comparing the two examples shows that the image quality has explicit consequences to the detection rate (72% vs. 78%).

<table>
<thead>
<tr>
<th>Total</th>
<th>Camera 1</th>
<th>Camera 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traffic Signs</td>
<td>306</td>
<td>160</td>
</tr>
<tr>
<td>Detected Traffic Signs</td>
<td>227 (74%)</td>
<td>118 (74%)</td>
</tr>
<tr>
<td>Not Detected Traffic Signs</td>
<td>79 (26%)</td>
<td>42 (26%)</td>
</tr>
<tr>
<td>False Detections</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

**Table 3: Results (all image sequences together).**

Table 3 shows the detection rate referred to the entire benchmark test (all two image sequences together) – detection rate is around 74% – sum of false detection is 5.

An additional example is shown in Figure 7 representing a small part of an image sequence including the detected SIFT feature points (red crosses) and the traffic signs (yellow rectangle).
Generally, the detection system works in a more than satisfactory way – more than 70% of the traffic signs can correctly be detected and classified. This value is a comparable value to other techniques and methods for traffic sign detection, e.g. Fang et al. (2003) or Oh (2005). 70% of detection rate seems not substantive, but has to be seen under the constraint that traffic signs are small objects in complex scenes.

The result of detection highly depends on the models, included into the system database and on the quality of the image sequences. Very impressive is the very low number of false detected objects. 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 ideal light conditions. Poor image quality can be improved by means of pre-processing (histogram equalization and median filtering should give adequate results).

As mentioned before one of the advantages of the presented approach is the runtime of the implemented algorithms. Currently, the whole process is implemented as a non-optimized version under different programming and script languages. Processing one image pair needs
about 2 seconds (Intel Core2Duo – 2 GB RAM). Optimizing the procedure by implementing the algorithms in C++/C#, using the Graphical Processing Unit (GPU) and by an intensive utilization of the multi-core capability of modern processing units could reduce runtime by a factor of 10.

5. DISCUSSION AND CONCLUSION

The article presents an automated traffic sign detection method which has been tested on image sequences captured by a mobile mapping system (VISAT™). 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 easily and the possibility to realize the whole procedure as real-time process.

A problem of the presented approach is that the SIFT algorithm employs scale-space technique and needs therefore images of high resolution to work in a satisfactory way. If the object of interest is very small, it is very difficult to detect the maximum/minimum from the scale space. The high resolution images captured by the used MMS are well suited for such an approach – problem may happen if the image is very small and blurred. Basically, a traffic sign is described by a low number of key-points (for the shown examples averagely about 15 key-points). Therefore, the image quality is a very important issue for the presented approach.

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, the integration of a Kalman Filter for predicting traffic sign location is envisaged.
REFERENCES


BIOGRAPHICAL NOTES

**Dr. Alexander Reiterer**
2001: Graduated Engineer, Vienna University of Technology, Austria.
2004: PhD at Vienna University of Technology, Austria.
2006 and 2008 Guest Reseracher at ETH Zurich and University of Calgary (Canada).
Currently: PosDoc at the Institute of Geodesy and Geophysics, Vienna University of Technology, Austria.

**Taher Hassan**
2002: Msc from Ain Shams University, Cairo, Egypt.
2009: PhD at University of Calgary, Canada.
Currently: Member of the MMSS Research Group, University of Calgary, Canada.

**Dr. Naser El-Sheimy**
1984: B.Sc. at ASU, Egypt
1989: Diploma at ITC, The Netherlands
1990: M.Sc. at ASU, Egypt
1996: Ph.D. at University of Calgary, Canada
Currently: Head of the MMSS Research Group, University of Calgary, Canada.

CONTACTS

**Dr. Alexander Reiterer**
Institute of Geodesy and Geophysics, Vienna University of Technology
Gusshausstrasse 27-29
1040 Vienna
AUSTRIA
Tel. +49-1-58801-12845
Fax +49-1-58801-12894
Email: alexander.reiterer@tuwien.ac.at
Web site: http://info.tuwien.ac.at/ingeo

**Taher Hassan**
Department of Geomatics Engineering, The University of Calgary
2500 University Drive, N.W
Calgary, AB, Canada T2N 1N4
Tel. +01-403-220-8149
Fax +01-403-284-1980
Email: tfabbas@ucalgary.ca
Web site: http://mms.geomatics.ucalgary.ca
Dr. Naser El-Sheimy
Department of Geomatics Engineering, The University of Calgary
2500 University Drive, N.W
Calgary, AB, Canada T2N 1N4
Tel. +01-403-220-7587
Fax +01-403-284-1980
Email: elsheimy@ucalgary.ca
Web site: http://mms.geomatics.ucalgary.ca