Visual Object Detection for Mobile Road Sign Inventory[1]

Christin Seifert[1], Lucas Paletta[1], Andreas Jeitler[2],
Evelyn Hödl[2], Jean-P. Andreu[1], Patrick Luley[1] and Alexander Almer[1]

[1]JOANNEUM RESEARCH Forschungsgesellschaft
Institute of Digital Image Processing
Wastiangasse 6, A-8010 Graz, Austria
E-mail: lucas.paletta@joanneum.at

[2]FH Joanneum GmbH
University of Applied Sciences
Alte Poststrasse 149, A-8020 Graz, Austria
E-mail: andreas.jeitler@fh-joanneum.at

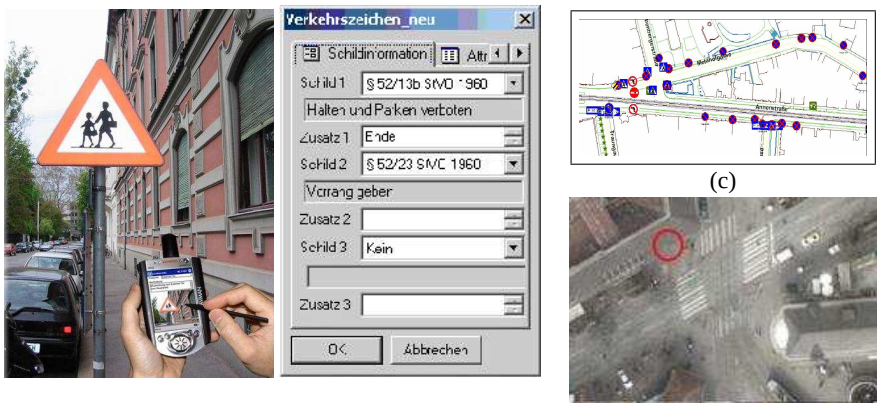
Abstract

For road sign inventory and maintenance, we propose to use a mobile system based on a handheld device, GPS sensor, a camera, and a standard mobile GIS software. Camera images are then analysed via object recognition algorithms which results in an automated detection, i.e., localisation and classification of the signs. We present here the localisation of points and regions of interest, the fitting of geometrical constraints to the extracted set of interest points, and the matching of content information from the visual information within the sign plate. From the preliminary operational state of the vision based road sign detection system we conclude that the selected methodology is efficient enough to achieve the requested high quality in object detection and classification.

1 Introduction

Road sign inventory and maintenance is today performed on the basis of GIS (geographic information systems) relying on geo-reference and content information about the signs. For the registration of previously undocumented sign objects, we propose to use a mobile system based on a handheld device, GPS sensor, a camera, and a standard mobile GIS software. The camera captures the appearance of the signs, images are then analysed via object recognition algorithms which results in an automated localisation and classification of the signs. Automated classification enables faster and more reliable processing than manual interaction with a GUI. We present here the full process chain for a robust recognition of sign objects: the localisation of regions of interest (ROI), the fitting of geometrical constraints to the ROIs, and the analysis of visual information within the sign plate. From the preliminary operational state of the mobile vision based inventory system we conclude that the selected methodology is efficient enough to achieve the requested high quality in object recognition.

[1]This work is funded by the European Commission's IST project DETECT under grant number IST-2001-32157, and the Austrian Joint Research Project 'Cognitive Vision', sub-projects S9103-N04 and S9104-N04.



(a) (b) (c) (d)

Figure 1: (a) Mobile inventory system, (b) manual sign definition interface, (c) sign symbols within ArcPad plan, (d) air-borne images for positioning.

2 Mobile Inventory System

The goal of the system is to attain efficient acquisition, attribution, and localisation of road signs for the purpose of inventory, inspection and maintenance. The system is based on ArcPad (CE Software) on a handheld device, GPS localisation, and a vision based module for object recognition. The mobile prototype was developed on a Pocket PC (HP's iPAQ 3970) in connection with a blue tooth GPS receiver (Fortuna BTGPS). A digital image of the road sign is captured by the mobile user within the urban/rural environment. The PDA based ArcPad application supports the data acquisition (Fig. 1a) and overtakes the central data storage. It allows automatic gathering of parameters like GPS position, current time stamp (UTC) and also the user ID. The acquisition of field data is assisted using a "moving map" functionality based on GPS location and accordingly transformed handheld position into a current map projection. The uncertainty in the positioning is compensated by a 'snap-to-point' method, in case the location of the road sign would be known beforehand. We will use standard auxiliary tools for geo-referencing, e.g., indication of location on maps or airborne image data, etc., Fig. 1d. It is planned to integrate the camera into the mobile device, and to outline a client based object detection module. System enhancements of 'mobile vision' capabilities will allow to automatically gather road sign attributes, which can be supplemented and also changed by the user, and will also perform validity and plausibility tests on the acquired data. For the detailed attribution of the collected data, we currently apply an off-site post-processing, consisting of, (i) geo-referencing of road signs via time synchronisation between digital image and track log data of the GPS receiver, and (ii) object detection of the automated identification of every road sign in the image.

(a) o_1 (b) o_2 (c) o_3 (d) Content

Figure 2: Road sign subwindow detections (a)-(c) and associated sign plate content (d) corresponding to objects in (a)-(c).

3 Visual Object Detection

The goal of the object detection module is to localise and identify road signs within the captured images. The images can be taken from a camera attached to the handheld, or a connected digital camera. The image should then be either transmitted to a server, or directly processed on the mobile client. While this work considers the first case, the latter is still topic of ongoing research.

The demand on the quality of service is high, in order to represent a beneficial technology for mobile road sign inventory and maintencance, and to substitute any manual interaction. Therefore, we aim at 100 $\%$ detection rate without positive falses. Efficient visual recognition of road signs has already been proposed in the context of automotive applications (Gavrila and Philomin, 1999), and been refined with respect to ambiguous information when recognising signs from a moving vehicle, e.g., in (Escalera et al., 2003). However, all these systems have in common, that (i) they were operated on images taken from a moving vehicle, under bad weather conditions, or just from a far distance, and accordingly, (ii) the recognition rate was 'clearly below' 100 $\%$. In the proposed application, (i) the mobile user has an impact on image acquisition (using flash and appropriate distance to object, (ii) we are required to achieve a detection rate above 98% to compete efficiency in manual interaction.

We follow here a framework of cascaded processing on the visual information, in order to become more robust in the interpretation, as follows,

1. **Pixel classification from learned color filters.** We learn color filters from images to classify pixels, applying an EM (expectation maximisation (Dempster et al., 1977)) cluster algorithm and a maximum likelihood classification approach thereafter.
2. **Local regions of interest.** Local subwindows (e.g., 10×10 pixels) are then interpreted for further processing (Fig, 3a,d). We exploit the information from

bimodal distributions of color pixels (in analogy to (Matas et al., 2000)), characterising typical color contrasts found in road signs, such as, 'RW' (red-white), 'BW' (blue-white), and 'RB'. We examine a plausability test by applying rules for window classification, such as, 'if W $> 35\%$ AND R $> 35\%$ then RW'.

3. **Extraction of sign geometry.** We then extract triangle and circular structure from local subwindows. We apply two primitive extractors, (i) an ellipse fitter (Fitzgibbon and Fisher, 1999) and (ii) Hough Transform (Illingworth and Kittler, 1988) to extract straight lines (Fig. 3b,e). The primitives are determined either on center points of the local subwindows. We perform some post-processing, such as, clustering similar lines, rejecting 'useless' lines and ellipses, i.e., those who have not sufficient support from the detected subwindows.
4. **Matching the sign content with prototypical patterns.** The final step is to extract the sign content, and apply a matching (using correlation) to stored prototypical road sign patterns, for road sign identification (Fig. 3c).

The experiments demonstrate that the preliminary system already achieves very robust interpretation of the road sign images. We intend to (i) extend the object database to classify up to 120 Austrian road signs, and to (ii) make the approach more robust by applying a probabilistic framework all over the recognition process. 'Mobile Vision' is an emergent technology in mobile computing, with potential applications in automated translation (Yang et al., 2001) and object based tourist information systems (Fritz et al., 2004).

References

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B*, 39(1):1–38.

Escalera, A., Armingol, J., and Mata, M. (2003). Traffic sign recognition and analysis for intelligent vehicles. *Image and Vision Computing*, 11(3):247–258.

Fitzgibbon, A. W.and Pilu, M. and Fisher, R. B. (1999). Direct least-squares fitting of ellipses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(5):476–480.

Fritz, G., Seifert, C., Paletta, L., Luley, P., and Almer, A. (2004). Mobile vision for tourist information systems in urban environments. In *International Conference on Mobile Learning, MLEARN 2004*.

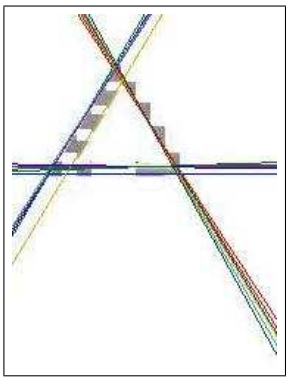
Gavrila, D. and Philomin, V. (1999). Real-time object detection for smart vehicles. In *Proc. International Conference on Computer Vision*, pages 87 – 93.

Illingworth, J. and Kittler, J. (1988). A survey of the Hough transform. *Computer Vision, Graphics, and Image Processing*, 44:87–116.

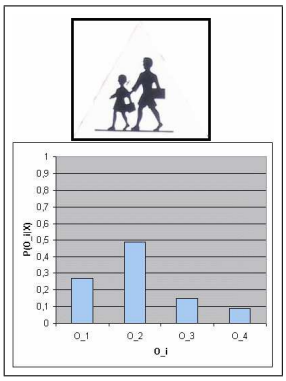
Matas, J., Koubaroulis, D., and Kittler, J. (2000). Colour image retrieval and object recognition using the multimodal neighbourhood signature. In *Proc. European Conference on Computer Vision*.

Yang, J., Gao, J., Zhang, Y., Chen, X., and Waibel, A. (2001). An automatic sign recognition and translation system. In *Workshop on Perceptive User Interfaces*. ACM Digital Library.

(a) Subwindows

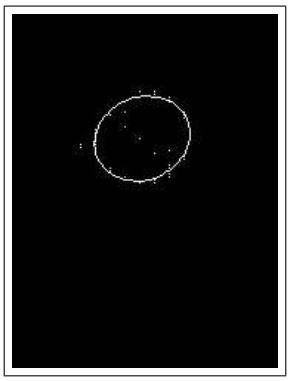


(b) Hough transform



(c) MAP Object hypothesis (o_2)

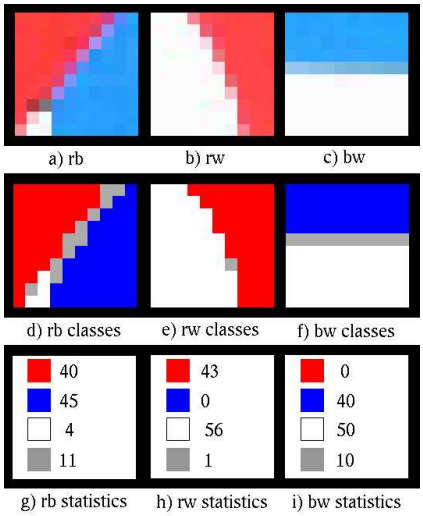
(d) Subwindows



(e) Ellipse fitted

(f) Ellipse overlaid

(g) Selected subwindows



(h) Subwindow content: a.c) original colors, d-f) color classification, g-i) class statistics

Figure 3: Cascaded road sign detection.