VISUAL RECOGNITION OF BRIDGES BY USING STEREO CAMERAS ON TRAINS

E. Funk[1,2], L. S. Dooley[2], S. Zuev[1], A. Boerner[1]

[1]German Aerospace Center (DLR)
Institute of Robotics and Mechatronics,
Berlin, Germany

{eugen.funk, sergey.zuev, anko.boerner}@dlr.de

[2]Department of Communication and Systems,
The Open University, Milton Keynes, MK6 7AA
United Kingdom

l.s.dooley@open.ac.uk

ABSTRACT

Recognition of either patterns or objects in mobile systems continues to be in the focus of intensive research, with many applications being enhanced by integrating environment related information. This paper presents a practical technique for detecting and recognizing bridges from a train using a stereo camera which provides depth and grayscale images. The algorithm has been applied to a train system, where object detection combined with a given map of an area is used to improve localization. The approach is based on the detection of primitive features including edges and corners in the depth image. The pairwise spatial relations between the features are then modeled by a graph, so the classification and detection can be performed by a probabilistic Markov Random Field framework. The algorithm has been tested on the real-life datasets of the *Rail Collision Avoidance System (RCAS)* project. The presented results prove the applicability of the framework for detection of objects by exploiting geometrical appearance constraints.

1. INTRODUCTION

Object recognition for transportation systems has recently received considerable attention by the research community (Quddus, 2008; Choi, 2010) with modern trains and railway supporting systems being required to detect either people or obstacles on the tracks. In both scenarios the object detection and recognition process is intractable using common cameras. In this work, the Integrated Positioning System (IPS) (Grießbach et al, 2010) has been used for visual perception, which has been developed for indoor and outdoor navigation using a stereo camera and *Inertial Measurement Unit* (IMU) sensors. The stereo images are processed by a disparity-matching algorithm (Hirschmüller, 2005) and the resulting depth image provides a pixel-wise distance to observed objects relative to the camera.

The remainder of the paper is organised as follows: Within the next section, related work is being reviewed including methods to detect bridges and tunnels in stereo sequences, while Section 3 introduces the IPS system and describes the concepts behind the new detection algorithm. Section 4 presents some detection results, obtained on real-world data. Section 5 provides some concluding comments.

2. Related Work

Three key areas of object detection and its application on train systems include: i) safety relevant applications (people and obstacle detection on rail tracks (Oh et al, 2010), (Rüder et al, 2003), ii) map building and refining (Gui-gui, 2006) and iii) navigation using an existing map and the detected objects in the scene (Rahmig et al, 2012). The object detection approach presented in this paper applies the idea of part-based object detection (Felzenszwalb, 2010). Similarly, directed Gabor-filters (Feichtinger, 1998) are used to detect salient edges in the depth images. In contrast to the Felzenszwalb's approach, the filter responses are clustered into connected groups in order to reduce the number of edge points. The spatial edges in the image are then used in a probabilistic Markov Random Field (MRF) (Koller, 2011) framework, to enable the detection and classification of objects in the scene by regarding the geometrical structure of the edge groups. This applied MRF approach is similar to (Qian, 1997), who has previously used it for detecting and classifying faces based upon independent eye, mouth and nose observations.

3. The System Concept

The system architecture shown in Figure 1 consists of three basic processes: feature extraction, region-based clustering and the main contribution of this work, the MRF framework for structural object detection and recognition. Each of these process modules will now be individually described.

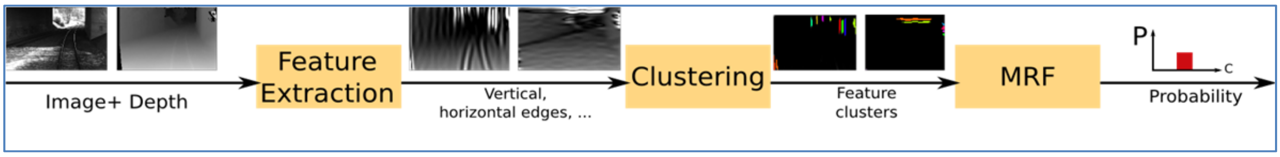


Figure 1: Architecture of the Detection System.

3.1 Feature Extraction

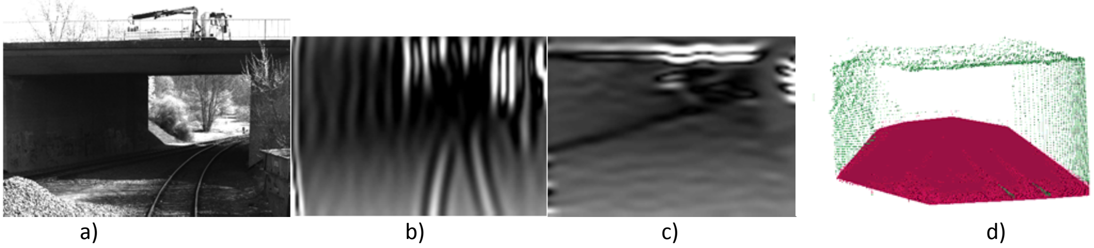


Figure 2: a) The original image from the left camera. b, c) Gabor-Filter results for detection of vertical and horizontal edges. d) The detected ground plane (solid) with 3D points reconstructed from the stereo images.

The feature extraction process deals with the detection of simple salient points in the image that correspond to edges and corners in three dimensions. To achieve this, a convolution filter is applied to the depth image to give an edge intensity for each pixel. The detection of edges is performed by a Gabor Filter (Feichtinger, 1998), which calculates the gradient in arbitrary directions. Figure 2 illustrates examples for vertical and horizontal Gabor-Filters.

The detection of spatial corners is achieved by extending the Harris Corner Detector (Harris, 1988), which calculates the covariance matrix for every 3D point using its local neighbours (Funk, 2011). After converting every pixel from the depth image to a 3D point using the geometrical stereo information of the IPS system, the 3D corner detector is applied in the same manner as for images, calculating the corner intensity for every point as shown in Figure 2b).

In addition to the presented convolution filter methods, the ground plane of a scene is being estimated and considered as a feature. Using the *random sample consensus* (RanSaC) (Fischler, 1981) approach the most significant plane in the scene is identified by selecting three random points and then calculating the spanning plane and the number of intersected points from the scene. Figure 2d) shows the corresponding result.

Having defined a small set of simple features, the extraction process is computational inexpensive, though a high number of extracted corners and edges will be redundant and so must be reduced to groups of either similar edges or corners. To achieve this, clustering techniques are adopted to reduce the number of detected corners.

3.2 Region Based Clustering

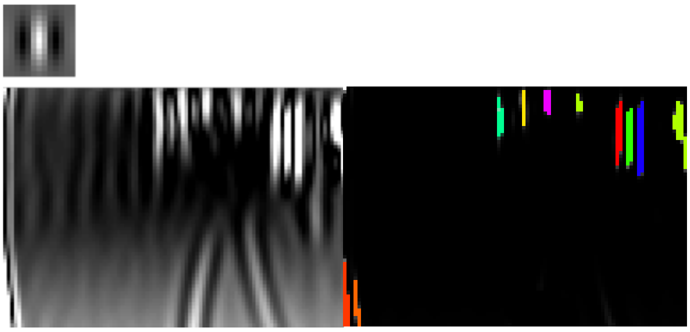


Figure 3: Grouping the feature points into connected sets.

Clustering aims at reducing the number of features. As the corner filters generate intensity values for each pixel, this means there are many feature points to be compressed for the MRF framework presented in the next section.

A region-based clustering approach (Figure 3) is applied to a set of points with a predefined measure of similarity. The similarity is defined as the Euclidean distance between two pixels and if this is less than a predefined threshold then the two pixels are merged. The benefit of this approach is its flexibility as it does not require the user to define the number of clusters as in classical clustering techniques such as the *k*-means (MacQueen, 1967) algorithm.

3.3 The MRF Framework

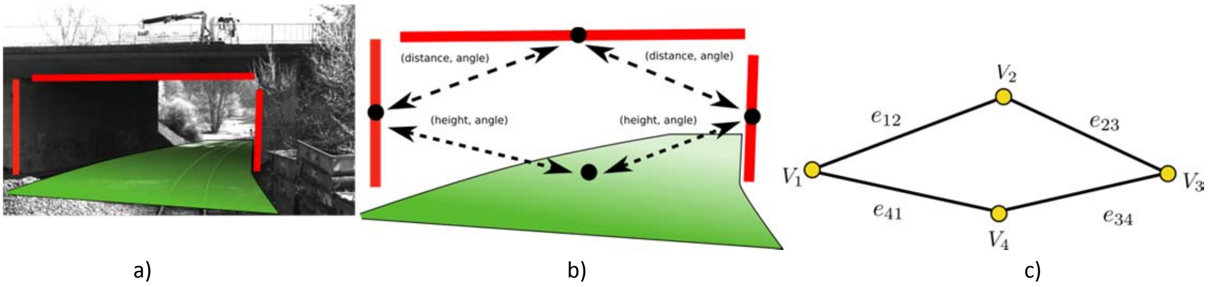


Figure 4: a) Four basic features define the bridge model. b) Pairwise relations between the features. c) Abstracted graph for the bridge.

In this work, the MRF framework (Koller, 2009) is used to model geometrical objects by a small set of basic features. In contrast to dictionary-based methods (Jiang, 2011), this approach enables the integration of evidence concerning the spatial relations between the features. As illustrated in Figure 4a), a bridge can be modeled by four different features: two vertical edges, one horizontal edge and the ground plane. The model restricts the configuration (Figure 4b)) of the basic features and penalizes the distances and angles between the two features which are out of bounds. Using a graph to describe the pairwise relationship between the features, the MRF framework can be applied. The MRF is formally defined by the probability density function:

$$p(x|y) = \frac{\prod_C e^{-U(x_c)}}{Z}, Z = \sum\nolimits_X \prod_C e^{-U(x_c)} \qquad \textbf{(1)}$$

where x represents the selected set of features, y is the considered object class by comparing the model with the selection x and Z is the *partition function* which upholds:

$$\int_X p(x|y) = 1 \qquad \textbf{(2)}$$

The *potential function* $U(x_c)$ evaluates the pairwise compatibility of the given features in x. An important property of MRF is that the probability of the graph in Figure 4c) can be calculated by using only the potentials of the pairs, called cliques. Thus, the probability that the selected features x actually represent a bridge can be determined from:

$$p(x|y) = \frac{e^{-U_y}}{Z}, U_y = \phi_{y12}(x_1, x_2)\phi_{y23}(x_2, x_3)\phi_{y34}(x_3, x_4)\phi_{y41}(x_4, x_1) \qquad \textbf{(3)}$$

The pairwise potentials (ϕ_{y12} ... ϕ_{y41}) compare the pairwise relationships of the two selected features of the model configuration. For instance, if the first node is too far away from the second, the potential ϕ_{y12}will have a higher value and prevent (1) from returning a high probability value.
Having applied the MRF formalism to object detection the next task is to estimate the feature set x, which maximizes the probability of being part of object type y:

$$\hat{x} = \max_x p(x|y) = \frac{\prod_C e^{-U(x_c)}}{Z} \qquad \textbf{(4)}$$

This is a NP-hard problem so the method is restricted to only small sets of features. The order of time complexity then becomes:

$$O\left(\prod_F \binom{k_f}{n_f}\right) \qquad \textbf{(5)}$$

where k_f is the number of features of type f in the model y, n_f is the number of detected features in the scene, and the term $\binom{k_f}{n_f}$ is the binomial coefficient. Considering, for example, a scene having 20 vertical and horizontal edges: For the case of the bridge model, where two vertical edges, one horizontal edge and one ground plane, are required, the corresponding number of combinations to solve (4) will then be:

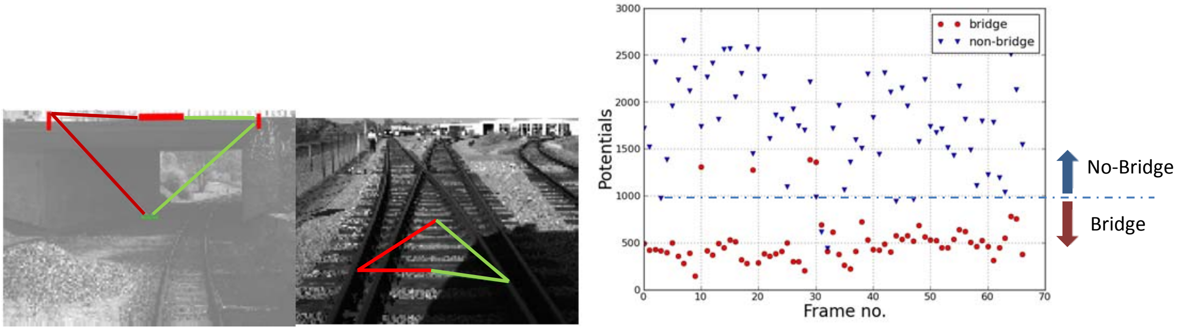
$$\binom{2}{20} \times 20 \ \times 1 = \ \ 3800 \,. \qquad \textbf{(6)}$$

Since $\hat{x}$ can only be a selection from features in one image, the estimation of the probability for $\hat{x}$ being of class y in general cannot be performed. In order to do so, the result of the potential calculation $\prod_C \, e^{-U(x_c)}$ from (4) must be classified by a training framework similar to the Support Vector Machine (SVM). However, in a simple case it is required to decide between two cases: Bridge is present or not present. Thus, a simple decision threshold for the potential result $U = \sum_c \, U(x_c)$ can be deduced from many processed scenes as illustrated in Figure 5c). A bridge is considered as detected when the unnormalized potential $\boldsymbol{U}$ exceeds a predefined threshold, illustrated as a separating line in Figure 5c).

4. Results

The novel detection algorithm has been evaluated with real-world datasets, obtained from the *Rail Collision Avoidance System (RCAS)* project.

Figure 5 presents the results for the detection of a bridge applied to a positive and negative scene. Figure 5c) illustrates the unnormalized potential value for U, which can be used for the classification of the detection result. The presented values of U have been obtained during a sequence of 60 frames, where every frame has been evaluated by the presented algorithm independently. U significantly increases in the no-bridge scenario. Thus, a systematic method for the design of MRF models needs to be considered. Figures 5a) and b) illustrate the configuration of the primitive features (edges and ground plane) for a given scene with the lowest potential value U.



a) b) c)

Figure 5: Bridge detection in a positive (a) and a negative (b) scene. The decision if a bridge is present can be made by the unnormalized potential values, illustrated in c).

5. Conclusion

An algorithm for the detection and classification of objects for trains has been presented. The approach utilizes local image features and the ground plane for the MRF framework for global matching and recognition. Since the recognition is performed on depth images provided by the IPS, the information about the position of detected objects is directly available. Further work will focus on the design of more salient local feature extraction methods and the systematic design of MRF models, which provide more unambiguous potential values resulting in more robust classification. Furthermore, the MRF framework will be extended towards the detection of objects with multiple occurrences.

References

(Rahmig et al, 2012) C. Rahmig, K. Lüddecke, K.Lemmer: Tunnels and Bridges as Observable Landmarks within a Modified Multi-Hypothesis Based Map-Matching Algorithm for Train Positioning. In: European Navigation Conference 2012, Gdansk, Poland, April 25-27, 2012.

(Quddus, 2008) Mohammed A. Quddus, Washington Y. Ochieng & Hongchao Liu. EDITORIAL: Special Issue: Intelligent Vehicle Navigation (Part1). Journal of Intelligent Transportation Systems 12:4, pages 157-158.

(Grießbach, 2010) D. Grießbach, A. Börner, I. Ernst, S. Zuev: Real-time dense stereo mapping for multi-sensor navigation. In: International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXVIII (5), pages 256-261. ISPRS 2010. ISPRS, Commission V, Midterm symposium, Newcastle 2010, Newcastle, UK.

(Choi, 2010) Kyoung-Ho Choi, Soon-Young Park, Seong-Hoon Kim, Ki-Sung Lee, Jeong-Ho Park, Seong-Ik Cho & Jong-Hyun Park, (2010) Methods to Detect Road Features for Video-Based In-Vehicle Navigation Systems. Journal of Intelligent Transportation Systems 14:1, pages 13-26.

(Oh et al, 2010) Seh-Chan Oh, Gil-Dong Kim, Woo-Tae Jeong, Young-Tae Park: Vision-based Object Detection for Passenger's Safety in Railway Platform. In: International Conference on Control, Automation and Systems 2008, Seoul, Korea.

(Rüder et al, 2003) M. Rüder, N.Möhler, F. Ahmed: An Obstacle Detection System for Automated Trains. Intelligent Vehicles Symposium, 2003. Proceedings. IEEE , vol., no., pp. 180- 185, 9-11 June 2003.

(Gui-gui 2006) Gao Gui-gui, Cai Bai-gen: Research on the Automatic Electronic Map Generation Algorithm for the Train Supervision System, In: Journal of the China Railsway Society, 2006.

(Hirschmüller, 2005) H. Hirschmüller: Stereo Processing by Semiglobal Matching and Mutual Information, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 328-341, February, 2008.

(Felzenszwalb, 2010) P. Felzenszwalb, D. McAllester, D.Ramann: A Discriminatively Trained, Multi-scale, Deformable Part Model, In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008.

(Feichtinger, 1998) Hans G. Feichtinger, Thomas Strohmer: Gabor Analysis and Algorithms, Birkhäuser, 1998.

(Qian, 1997) R. J. Qian: Object detection using hierarchical MRF and MAP estimation. In: Computer Vision and Pattern Recognition Conference, 1997.

(Koller, 2009) D. Koller: Probabilistic Graphical Models: Principles and Techniques, MIT Press, 2009.

(Harris, 1988), C. Harris and M. Stephens: A combined corner and edge detector, Proceedings of the 4th Alvey Vision Conference. pp. 147–151.

(Funk et al, 2011) E. Funk, D. Grießbach, D. Baumbach, I.Ernst, A. Boerner, S. Zuev: Segmentation of Large Point-Clouds Using Recursive Local PCA. In International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2011.

(Fischler, 1981) Martin A. Fischler und Robert C. Bolles: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, 1981.

(MacQueen, 1067) J. B. MacQueen: Some Methods for classification and Analysis of Multivariate Observations. In: Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. 1, University of California Press, 1967, S. 281–297.

(Jiang, 2011) Zhuolin Jiang, Zhe Lin, Larry S. Davis: Learning a Discriminative Dictionary for Sparse Coding via Label Consistent K-SVD. IEEE Conference on Computer Vision and Pattern Recognition, 2011.