A Novel Approach for Real-Time Semantic Context Labeling

Fons van der Sommen, Martin Pieck, Svitlana Zinger, Peter H.N. de With
Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven
{F.v.d.Sommen, M.A.R.Pieck, S.Zinger, P.H.N.de.With}@tue.nl

Abstract

The use of context information is an important aid for full semantic scene understanding in security and surveillance applications. This paper presents an innovative semantic context-labeling algorithm for three context classes – water, sky and vegetation – while trading-off quality and real-time execution. Our system consists of three consecutive stages: image segmentation, region-based feature extraction and classification. We propose the joint use of color features in HSV space, texture from Gabor filters and spatial context, in combination with the Directional Nearest Neighbor (DNN) method for segmentation. Compared to recent literature, this combination is over 35 times faster and achieves a coverability rate that is 65% higher [6].

1. Introduction

Semantic interpretation of image content is a challenging problem in computer vision. An algorithm for this type of image analysis is of high importance to the security and surveillance industry, because the manual analysis of video feeds is a tedious and error-prone process. Early approaches by Bao *et al.* [1] and Rankin *et al.* [7] focus on one context class and are extended too multi-class by Javanbakhti [2]. Based on the framework of Javanbakhti, we propose a near real-time context detection algorithm with an improved detection quality. The contributions of this paper are threefold. First, innovative methods are proposed for segmentation, texture features and classification, to optimize for both quality and execution time. Second, an improved quality performance evaluation is proposed and third, an efficient C++ implementation using multi-core functionality is presented that achieves near real-time execution.

2. Proposed method

The proposed framework is inspired by a recent region-based approach [2] and consists of three major stages: (1) segmentation of the image into regions, (2) the computation of region features and (3) the classification of each region.

2.1. Image segmentation

For segmentation of the image, variations of the graph-based segmentation algorithm by Felzenswalb *et al.* [4] are applied of which later a specific version will be chosen to embed in the framework. For construction of the graph we explore three options: Grid-based (8N), Neighbor-based (24N), as employed by Javanbakhti [2] and Directional Nearest Neighbor (DNN), as proposed by Liu *et al.* [5].

2.2. Feature extraction

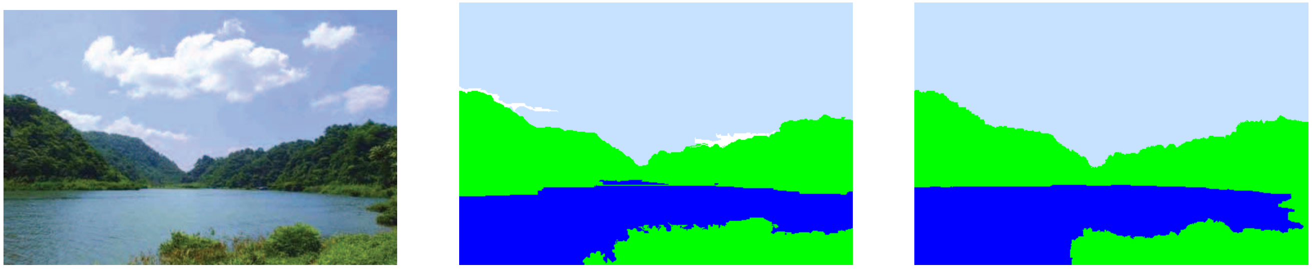
In the feature extraction stage, the HSV color intensity, Gabor-based texture information and Spatial Context (SC) are extracted from the regions resulting from the segmentation. In contrast to [2], we average features over the region instead of forming a feature vector for each pixel. This results in a lower computational complexity for classification. Also, we have analyzed the frequency spectrum of each semantic class to determine the optimal Gabor filter parameters, as proposed in [9]. Thus our approach attempts to reduce the number of image descriptors, while increasing their distinctive properties to preserve the quality.

2.3. Region Classification

In the last stage, every region is classified and labeled by a Support Vector Machine (SVM). The SVM is trained offline using a radial basis function kernel with 10-fold cross-validation combined with a grid search for optimal parameters C and γ. This search also incorporates the SVM model size, as fewer support vectors result in faster classification.

3. Experimental results

We evaluate the proposed semantic context-labeling algorithm on images of natural scenes, selected from the dataset used in [2], so that all images contain the semantic classes water, sky and vegetation. This set contains 54 images of which 49 have a 481×321 pixels resolution and 5 have a 320×240 resolution. The results regarding execution time are computed using only the 481×321 images to obtain comparable uniform results. For each image, the manually



(a) Original image. (b) Classified image. (c) Ground truth image.

Figure 1. A multi-class detection example. Sky is colored light blue, water is blue, vegetation is green and undetected regions are white.

Table 1. Comparison of quality performance for water detection.

	Features	Segm.	CR (%)	ER (%)
I	HSV	DNN	59.6	36.3
II	HSV+Gabor	DNN	79.3	19.1
III	HSV+Gabor+SC	DNN	92.2	12.9
IV	HSV+Gabor+SC	8N	84.0	14.0
V	Javanbakhti [2]	24N	26.1	16.2

Table 2. Time performance comparison for water detection in milliseconds for each algorithm stage: Pre-Processing (PP), Segmentation (SEG), Gabor Filtering (GF) and Classification (CL).

	PP	SEG	GF	CL	Total time
I	1	142	113	15	272
II	2	143 (paralel)		15	161
II	3	109 (paralel)		15	127
IV	1	106 (paralel)		18	125
V	1	587	58	3,754	4,400

annotated ground truth is available. Figure 1 shows an example image, including the corresponding ground truth and algorithm output. The SVM is trained on 24 images and the algorithm is tested using the remaining 30 images, where we measure the quality by the performance of Coverability Rate (CR) and Error Rate (ER) as proposed by Schmitt *et al.* [8]. The algorithm is implemented in C++, using OpenCV for image processing functions and the OpenMP API [3] for parallel functionality. The application is executed on an i7 quadcore @ 2.4-3.4 GHz processor with 8 GB DDR@1600 MHz RAM on 64-bit Windows 8.1 Pro.

Table 1 compares the quality performance for different segmentation and feature configurations. From the table, it can be observed that adding extra features for the same segmentation method (DNN) leads to higher performance. Namely, the CR increases while the ER decreases. The implementation with the highest performance, HSV+GABOR+SC in combination with DNN, outperforms the method of Javanbakhti by more than 65%.

Table 2 shows the results of timing analysis for each step in the algorithm. From the table it can be observed that by exploiting the parallelism, the total computation time can be reduced from 272 ms to 161 ms. Applying extra parallel functionalities, make our non-parallel approach more than 2.2 times faster, leading to overall 125-ms execution time.

4. Conclusion

In this paper, we propose a novel semantic context labeling algorithm that exploits the joint use of the HSV color features, Gabor filters and spatial context in combination with a DNN segmentation method. Experiments show that it is over 35 times faster and achieves a CR that is 65% higher, compared to state-of-the-art alternatives.

References

[1] X. Bao, S. Zinger, R. Wijnhoven, et al. Water region detection supporting ship identification in port surveillance. In *Advanced Concepts for Intelligent Vision Systems*, pages 444–454. Springer, 2012. 1

[2] I. M. Creusen, S. Javanbakhti, M. J. Loomans, L. B. Hazelhoff, N. Roubtsova, S. Zinger, et al. Vicomo: visual context modeling for scene understanding in video surveillance. *Journal of Electronic Imaging*, 22(4), 2013. 1, 2

[3] L. Dagum and R. Menon. Openmp: an industry standard api for shared-memory programming. *Computational Science & Engineering, IEEE*, 5(1):46–55, 1998. 2

[4] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004. 1

[5] Z. Liu, D. Hu, H. Shen, and G. Feng. Graph-based image segmentation using test directional nearest neighbor graph. *Science China Information Sciences*, 56(11):1–10, 2013. 1

[6] M. A. Pieck, F. van der Sommen, S. Zinger, E. J. Schoon, and P. H. de With. Real-time semantic context labeling for image understanding [to be published]. In *IEEE International Conference on Computer Vision (ICIP)*, 2015. 1

[7] A. L. Rankin, L. H. Matthies, and A. Huertas. Daytime water detection by fusing multiple cues for autonomous off-road navigation. Technical report, DTIC Document, 2004. 1

[8] F. Schmitt and L. Priese. Sky detection in csc-segmented color images. In *VISAPP (2)*, pages 101–106, 2009. 2

[9] F. Van Der Sommen, S. Zinger, E. J. Schoon, and P. H. de With. Supportive automatic annotation of early esophageal cancer using local gabor and color features. *Neurocomputing*, 144:92–106, 2014. 1