Integration of Regions and Contours for Object Recognition

D. Schlüter, F. Kummert, G. Sagerer
Bielefeld University
Faculty of Technology, Applied Computer Science
D-33594 Bielefeld, Germany
{dschluet,franz,sagerer}@techfak.uni-bielefeld.de

S. Posch
University of Halle–Wittenberg
Institute of Computer Science
D-06099 Halle/Saale, Germany
posch@informatik.uni-halle.de

Abstract

We present an integrated approach combining region and contour–based techniques to enhance both segmentation and recognition processes. This cue integration operates on the level of contour–based groups and complete regions, which are matched to reflect a common cause in the image (and thus in the scene). Additionally, we realize a top–down scheme controlling the segmentation process on the basis of unresolvable image areas and the integration of processes on different time scales using a region memory. Results for real images are given showing distinct improvements of recognition results.

1. Introduction

Region and contour–based techniques are well known as being dual approaches to image segmentation suffering from different drawbacks (and enjoying different advantages, as well). It is therefore attractive to combine both approaches to eliminate their drawbacks. While most work to implement such a cue integration relies on processing at the pixel level (see e.g. [4]), we propose to integrate both pathways at a more abstract level of representation, namely complete regions and contour segments derived from perceptual grouping. The results of this segmentation along with focus points from still another path of processing are the bases for a subsequent hybrid object recognition module in our system [1]. In addition we make use of intermediate results of object recognition to focus the segmentation and integration process to image areas with unidentified objects, thus realizing a top-down scheme to control processing (for other work for such a top-down control see e.g.[3]). When processing image sequences we have also to be able to cope with different time scales of the path ways, namely contour– and region–based processing. Therefore, we introduce a region memory making use of the stability of regions in the course of time and again matching of contour-segments and regions. We present results for real images taken from our domain of a toy construction kit.

2. Segmentation

Color-Based Region Segmentation

This segmentation process is based on the fact that in our domain only a fixed set of different colors may occur. The approach starts by classifying each single pixel of the image into one color class. For that, a polynomial classifier of sixth degree uses the YUV-values of a pixel as a three-dimensional feature vector. To speed up this time consuming process a look-up table is used. As the classification process uses only local information of a pixel some errors may occur due to noise. Therefore, a smoothing operation is applied deciding for the color with maximum occurrence within an 11 by 11 window.

Subsequently, a connected components analysis marks each pixel of a given region with one unique label. Finally, for every such a region the boundary is approximated by a polygon and appropriate features like compactness or eccentricity are calculated.

Contour-Based Perceptual Grouping

In this section we give an overview of the contour-based segmentation and grouping. The initial contour segmentation uses a standard Sobel operator and a subsequent thinning with non maximum suppression and hysteresis threshold. The resulting edge elements are approximated with straight line segments and elliptical arcs with the method due to Leonardis (see [8]).

These contour segments are subsequently organised within a hierarchy of grouping hypotheses (see Fig. 1) using various Gestalt laws. In the lowest level contour segments are combined to form collinear and curvilinear groups which are again approximated with straight line segments or elliptical arcs. These groups as well as the initial contour segments are denoted as *linear groups* subsequently. In addition, in this level pairs of linear groups may form a proximity grouping if their endpoints are spatially close.

In Proc. 15th International Conference on Pattern Recognition, volume I, pages 944-947, Barcelona, September 2000. IEEE.

The next 2×1D level deals with pairs of symmetric or parallel linear groups. To generate hypotheses within the 1D and 2×1D level of the hierarchy we employ the concept of *Areas of Perceptual Attentiveness* which are derived from a hand labelled training set (see [9]).

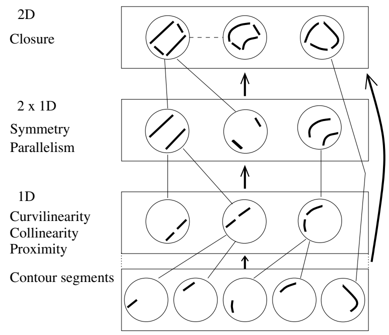

Fig. 1. Group hierarchy

In the highest level linear groups are organised into closed contours. To this end we construct an undirected graph with all linear groups. The endpoints of the groups represent the nodes in the graph, and the edges of the graph are distinguished into *contour edges*, modelling the underlying contours of a group, and *proximity edges*, connecting two linear groups at their endpoints due to a proximity group. This so-called *proximity graph* for a synthetic example is depicted in Fig. 2. In this graph a closed contour corresponds to a simple cycle with an alternating sequence of contour and proximity edges, which do not intersect.

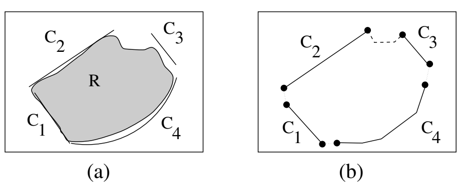

Fig. 2. (a) Synthetic region and contour segmentation; (b) Sample proximity graph

The generation of all grouping hypotheses within the hierarchy relies on local evidence of the image data only. Therefore, all hypotheses are subsequently judged in a more global context using a Markov Random Field. This results in a subset of significant groups yielding a globally consistent interpretation of the image data. For details on global judgement and the generation as well see [10, 11].

Matching of Contours and Regions

In some situations searching for closed paths of linear groups may fail, because a proximity group between two linear groups is missing. For example low contrast at the object boundary prevents a correct approximation or yields a wide gap. To overcome these situations we integrate the colorbased region information to support the grouping process by defining *matches* between groups and regions. In our approach a *match* describes a relation between a region and a contour group in the sense of being caused by the same event in the scene, e.g. the projection of the same surface of an object. For this the distance of the endpoints, and an approximated mean distance of the contour group to the region boundary are calculated. These features are used by a polynomial classifier to classify each group to approximate the *boundary* or describe internal *structure* of a region. Thus, the boundary match of two linear groups to the same region support the generation of a *region-based proximity* along the region boundary in the proximity graph. This region-based proximity is used to close a path and generate a closed contour (see Fig. 2: edge connecting C_2 and C_3, details in [13, 11]). Thus, the integration of linear groups and regions is utilised to enhance the grouping process. In the next section we describe how the matches are employed to enhance a region-based object recognition process.

3. Cue Integration for Object Recognition

Before we outline our integration approach in detail the hybrid object recognition module is briefly described.

Hybrid Object Recognition

For the recognition of parts of our construction-kit we combine the semantic network language ERNEST [7] and a special form of neural networks called Local-Linear-Map (LLM) [12]. On the basis of the segmented regions the neural network uses the centroids of mass as 'focus points'. At each focus point and based on an edge enhanced intensity image a feature vector is extracted by 16 Gabor filter kernels. This is the input for the LLM-network calculating up to three competing object hypotheses [2]. For each competing LLM-hypothesis a so-called holistic instance inside the semantic network is created which are stored in competing search tree nodes (see Fig. 3 for a network overview). Dependent on the object type detected by the LLM-network

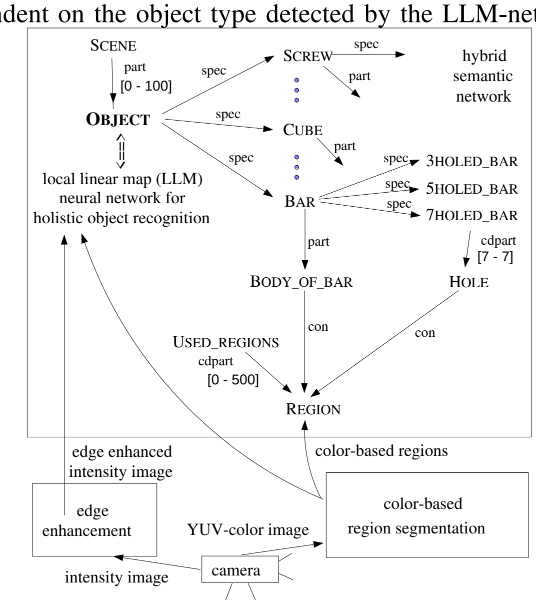

Fig. 3. The hybrid recognition system

an appropriate specialisation is selected to verify the object hypothesis according to the structural knowledge represented in the semantic network [5]. During this instantiation process restrictions for position, color, and shape are propagated in a model-driven way. In order to speed up the recognition process we extended this approach to image sequences where on the basis of previous results the current image is analysed [6]. Thus, temporally linked results are calculated supporting the recognition of events and actions.

Integrating Contour-Based Information

The object recognition process described in the previous section relies only on iconic image data like regions and focus points. This information is sufficient for robust object recognition in cases of differently colored objects. However there are several drawbacks inherent to region segmentation which can be resolved by combining it with contour information. We combine both cues at the level of judged contour groupings and complete regions, since contours and their groupings contain more abstract information about the global structure of an image than pixel based regions do.

A typical drawback of region segmentation occurs in situations where a projection of an object surface is fragmented due to occlusion by other objects: any region segmentation will produce disconnected regions for the bar in Fig. 4(a-c). The contour-based groupings help to resolve this drawback by generating linear groups for the fragmented contours of the bar (see Fig. 4(b)), and thus a closed contours for the correct surface projection of the bar is detected. Another situation occurs when two wooden bars overlap each other (see Fig. 4(d-f)). Region segmentation techniques inherently tend to generate one region for both bars, because the inhomogeneities at the boundary between both surfaces are usually insufficient to separate them. In this case the linear groups and closed contours describing the surfaces of the two bars support a separation of the region into the correct surface projections. Similar situations

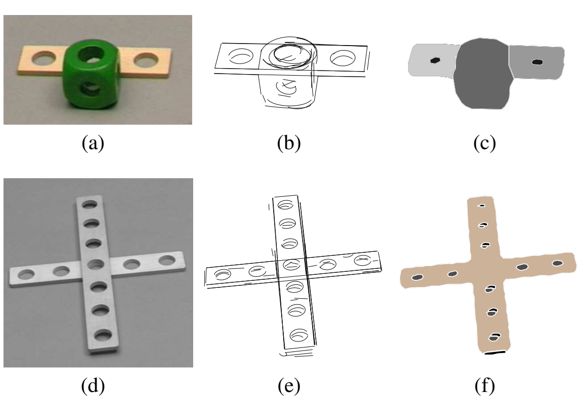

Fig. 4. Examples of contour-based grouping and region segmentation

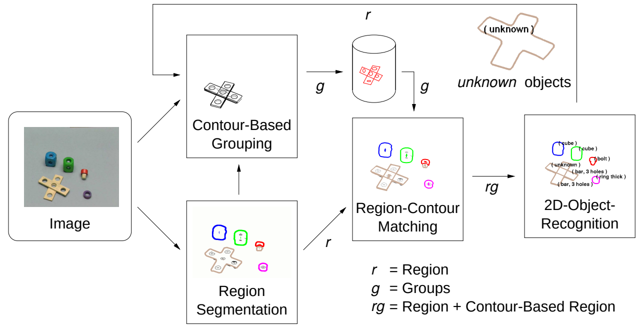

Fig. 5. Integration of contour grouping and region segmentation for object recognition

occur in more complex scenes as well.

Currently we use contour based groupings in a top-town fashion in situations which cannot be resolved by the object recognition process. To this end the object recognition process classifies unresolvable colorbased regions as *unknown* objects. These regions are used as regions of interest (ROI) for contour-based segmentation and grouping. Thus, the contour-based process focuses on areas where colorbased information is insufficient and more structural information is needed. The resulting significant contour groups are matched to the current colorbased regions and every matched closed contour is used to calculate region features (the color is determined from the matched region). These *contour-based regions* are added to the colorbased regions yielding a set of regions from different, independent image segmentation techniques. Out of this set the recognition process classifies regions as outlined in section 3, whereas the contour-based regions enhance the recognition on the basis of structural contour information in unresolvable areas.

A significant problem of this strategy arises from different runtimes of the modules when analysing image sequences. The ratio of runtime between colorbased region segmentation and contour-based grouping is about $1 : 15$. However we want to preserve a fast object recognition on image sequences with the frequency of region segmentation. To this end a region memory identifies identical regions over image sequences on the basis of region features [6]. Along with this identification, the region memory indicates the *stability* of a region by saving a time stamp of its last significant change. This stability is utilised by the matching module to insure that contour groups are combined with regions only, if region features have not changed during the grouping process. However with this strategy the contour groups support the recognition process only about every 15th frame with structural information. Therefore the groups are stored in the matching module and are only matched to regions of a newer image if the stability is sufficient. If new groups are completely generated for an image, they replace the former groups in the matching module and thus new structural information is incorporated.

4. Results

In this section we present results of our integration approach for real images and the enhancement of object recognition. Figure 6(a) shows a typical scene of our domain, a plane built with a wooden toy construction kit. The results of the colorbased region segmentation is depicted in Fig. 6(b). Obviously the several wooden bars are segmented to one region of wooden color yielding an unknown object by the object recognition process. This region is used as a goal directed ROI for perceptual contour grouping which generates contour-based regions for the bars forming the wings of the plane (see Fig. 6(c)). Thus, the object recognition process is improved classifying contour-based regions as additional correct objects (see Fig. 6(d), (e) with additional recognised objects from contour-based regions). Another example is shown in Fig. 7. Objects of the same color overlap each other (wooden bars) and some are placed close to each other (3 green cubes). Again, integrating the contour-based regions into the region-based object recognition yields correct decompositions of objects. In addition also incorrect and unknown object hypotheses are generated (see Fig. 7(b), (c)), resulting from contour-based regions inside the green regions of the cubes. These regions have similar features as a green bolt in our domain. Nevertheless, the presented results show a significant improvement of the object recognition process based on our integration approach.

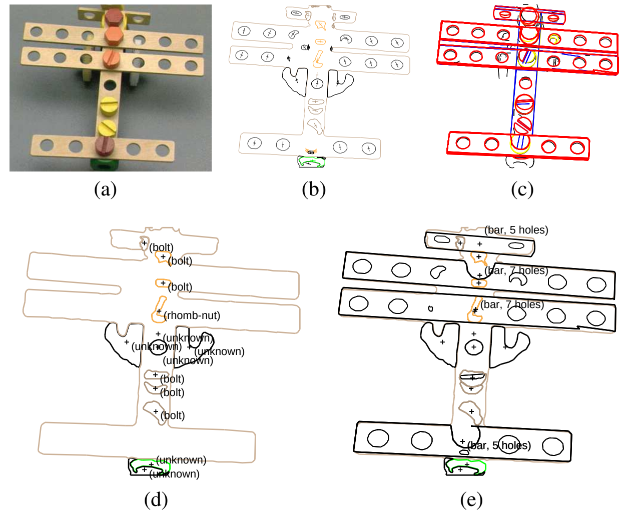

Fig. 6. Example of a plane

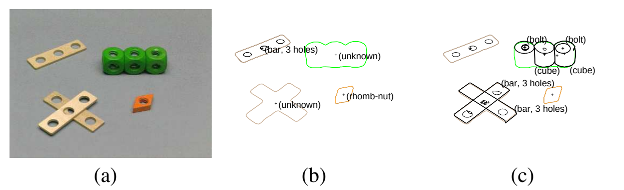

Fig. 7. Example of other objects

5. Conclusion

In this paper we propose to combine region-based and contour-based segmentation techniques for object recognition at a level, where both pathways interpret the image data on the level of complete regions and contour groups. With this combination and a top-down strategy including a region memory handling different time scales the object recognition process is significantly improved in situations where projections of 3D surfaces are merged due to homogeneous pixel values. Current work is concerned on extentions of the semantic network and the utilisation of closed contours to merge separated colorbased regions as shown in the top row of Fig. 4.

References

[1] C. Bauckhage, G. Fink, G. Heidemann, N. Jungclaus, F. Kummert, S. Posch, H. Ritter, G. Sagerer, and D. Schlüter. Towards an image understanding architecture for a situated artificial communicator. *International Journal of Pattern Recognition and Image Analysis*, 9(4):542–, 1999.

[2] G. Heidemann, F. Kummert, H. Ritter, and G. Sagerer. A Hybrid Object Recognition Architecture. In C. von der Malsburg, W. von Seelen, J. Vorbrüggen, and B. Sendhoff, editors, *Artificial Neural Networks – ICANN 96, 16.-19. July*, pages 305–310. Springer-Verlag, Berlin, 1996.

[3] C. A. Kohl. Goal-directed control for computer vision. COINS TR 88-22, Univ. of Massachusetts at Amherst, 1988.

[4] P. Kohlhepp. Combining region and edge based segmentation for robust 3d localization in range images. In *Proc. 2nd ACCV*, pages 254–258, Singapore, 1995. IEEE.

[5] F. Kummert. *Interpretation von Bild- und Sprachsignalen — Ein hybrider Ansatz*. Shaker Verlag, Aachen, 1998.

[6] F. Kummert, G. Fink, G. Sagerer, and E. Braun. Hybrid Object Recognition in Image Sequences. In *14th ICPR*, pages 1165–1170, Brisbane, Australia, 1998.

[7] F. Kummert, H. Niemann, R. Prechtel, and G. Sagerer. Control and Explanation in a Signal Understanding Environment. *Signal Processing*, 32:111–145, 1993.

[8] A. Leonardis. *Image Analysis Using Parametric Models*. PhD thesis, University of Ljubljana, 1993.

[9] A. Maßmann and S. Posch. Mask-oriented grouping operations in a contour-based approach. In *Proc. 2nd ACCV*, pages 58–61, Singapore, 1995. IEEE.

[10] A. Maßmann, S. Posch, G. Sagerer, and D. Schlüter. Using markov random fields for contour-based grouping. In *Proc. Int. Conf. on Image Processing*, pages 207–210. IEEE, 1997.

[11] S. Posch and D. Schlüter. Perceptual grouping using markov random fields and cue integration of contour and region information. TR 98/10, SFB 360, Univ. of Bielefeld, 1998.

[12] H. Ritter, T. Martinetz, and K. Schulten. *Neural Computation and Self-organizing Maps*. Addison-Wesley, Reading, MA, 1992.

[13] D. Schlüter and S. Posch. Combining contour and region information for perceptual grouping. In P. Levi, R.-J. Ahlers, F. May, and M. Schanz, editors, *Mustererkennung 1998.*, Informatik Aktuell, pages 393–401. Springer, 1998.