A multi-directional multiple-path recognition scheme for complex objects applied to the domain of a wooden toy kit ☆

2. Architecture

The proposed system architecture is character-ized by a combination of various processing paths (Fig. 1). The pathways deliver different kinds of results, that differ in their information content and also in their reliability and robustness. Moreover, the modules differ in the amount of task specific knowledge. The overall result profits from com-bining the complete and fast results of task specific modules with the robustness of more general ap-proaches.

The results of the recognition system, denoted by *object descriptions* in Fig. 1, are hypotheses for Baufix elements including information about the corresponding color region. In case of assemblies a description of their structure is additionally gen-erated.

Fig. 1. Outline of the system architecture and the processing paths.

One data driven processing path (1) is built up by a hybrid recognition module specialized for isolated Baufix elements (Heidemann et al., 1996). In assemblies often only parts of the Baufix ele-ments stay visible. Within the second pathway (2) small windows around interesting points in the image, in the following called (*classified*) *focus points*, are classified using a suitable set of classes for parts of Baufix elements (Heidemann, 1998). As an intermediate result between the object re-gions associated to the Baufix elements hypothe-sized within the first path and the point based results of the second path serves a color region segmentation using general color properties fol-lowing (Comaniciu and Meer, 1997) (3). These results form the data driven input of the central recognition unit. The assembly unit (4) generates model based descriptions of the structure of as-semblies built up from Baufix elements (Bauck-hage et al., 1998). It answers the question whether a recognition result leads to a possible assembly structure or not (5). The individual modules are described in more detail in Section 2.1.

In order to combine the different data driven results on the basis of the associated region in-formation, a voting scheme is applied. To get ad-ditional information, the central recognition tool marks supplementary focus points and connects the neural classifier (6). In case of competing in-terpretations, knowledge based and probabilistic reasoning within a Bayesian network is done. The structure for integration applied within the central recognition unit is the topic of Section 2.2.

2.1. Processing paths and modules

The first working path (1) provides hypotheses based on models for isolated Baufix elements (Heidemann et al., 1996).

For region segmentation, a pixel color classifi-cation uses a polynomial classifier to assign one of the Baufix colors to each pixel. Regions are es-tablished as areas of identical color class. The center of gravity of each region serves as a *point of interest*. For each such point, a 16-dimensional feature vector is extracted by 16 Gabor filters which are centered at the point of interest. These feature vectors are used by an artificial neural
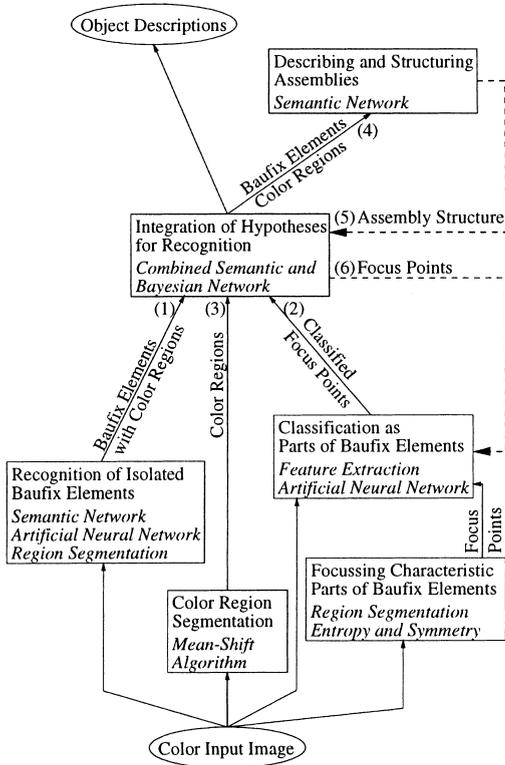
network for classification of isolated objects. These point based results initialize the instantiation process of a semantic network, that verifies them, exploiting knowledge about the fixed set of Baufix elements and the features of the color regions. The results are hypotheses for Baufix elements and the corresponding color regions. Because this module is optimized for the detection of isolated Baufix elements, the results are only reliable if elements are only slightly occluded (Fig. 2).

The second pathway (2) is designed for recognizing the occluded elements of complex objects by detecting characteristic parts of the elements (Heidemann, 1998).

Focusing these parts is done following two separate approaches. One way is to do a rough segmentation of the image by calculating the entropy map and mask those areas with high entropy or information content. Within these areas, for each pixel a symmetry value is calculated following the method of Reisfeld et al. (1995). In short, the symmetry measure judges the amount of symmetry of the gray value gradients within the surroundings of a pixel. From the symmetry values of each pixel a symmetry map arises, the maxima of which are taken as focus points. The second way of detecting interesting points is by the segmentation of the image by the color structure code (CSC) developed in (Priese and Rehrmann, 1993). This algorithm detects areas of homogeneous color. The center of gravity of those regions with high contrast at their boundary are taken as interesting points. Fig. 3 shows the focus points in an example image. The focusing modules follow general ideas of detecting interesting areas without modeling task specific knowledge. The adaption to the domain is done
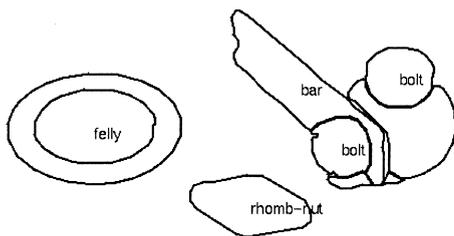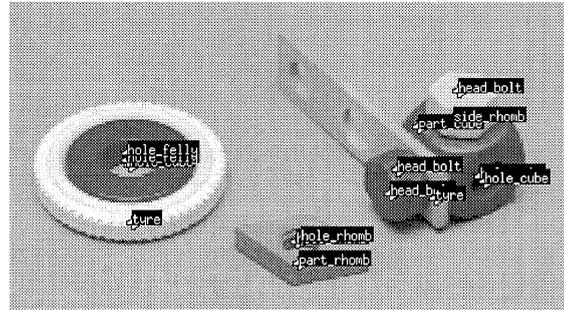

Fig. 3. Example of an input image showing the Baufix scenario with data driven focus points and classification results.

with only few parameters, specifying mainly the scale of interest and color contrast. After focusing a point in the image, a window around this point has to be classified as a part of a Baufix element. A set of labeled data is used to train the feature extraction and the neural classifier. The feature extraction consists of a vector quantization with a subsequent local principal component analysis for each reference vector. The feature vectors are classified by neural networks of the type Local Linear Map. Fig. 3 shows an example for the data driven focus points and their classification.

We use the Mean-Shift algorithm and pixel colors within the Luv color space for segmentation (3) (Comaniciu and Meer, 1997). The algorithm first reduces the number of colors occurring in the image by clustering within the color space. The number of clusters is not required as a parameter, but the threshold for the color distance is deduced from a segmentation parameter and the global variance matrix of the image. After pixels are associated to the color classes, this result is optimized by eliminating such color classes that only occur within small regions and finally re-sorting those pixels of very small regions to their spatial neighboring classes. This kind of region information uses general color properties and mediates between the object region hypothesized by the recognition module for the isolated elements and the point based information of the focus points. Fig. 4 shows an example of the segmentation results.

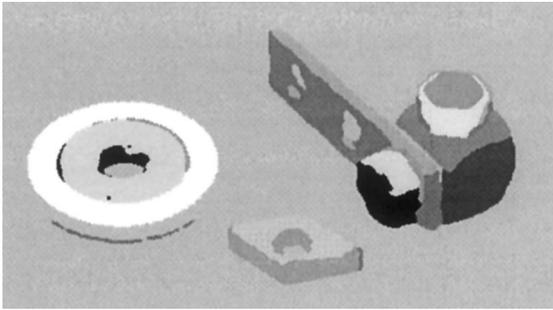For describing assembled objects, never seen before, a semantic network models all possible

Fig. 2. Result from the recognition unit specialized for isolated objects.

Fig. 4. Results of Mean-Shift color segmentation.

construction rules for the Baufix domain (Bauck-hage et al., 1998). Given a set of Baufix elements attributed by an estimation of their localization and geometry, the module decides whether this set can form an assembly and creates a suitable structure description. Exploiting the knowledge of assemblage, the module hypothesizes missing elements necessary for assemblage and gives hints to recognition failures.

2.2. Structure for integration

The central recognition unit integrates the different information sources by connecting the data driven and the model based attempts within a semantic network analysis of which the knowledge base is shown in Fig. 5.

A *bunch* of Baufix elements can contain one single Baufix element or several of them which lie next to each other or are assembled. On the more concrete layer a bunch of Baufix elements is rep-

resented by a number of neighboring color regions that represent one or more Baufix elements.

Starting from a color region that is not included within others, a Baufix element is hypothesized. In order to specialize this general hypothesis object information directly associated with the color region and information about parts of this Baufix element are considered. The *combination of different hypotheses* within the nodes is done by applying a plurality *voting scheme* (Parhami, 1994).

Parts are hypothesized by exploiting the classification results for the focus points that are geometrically associated to regions included within the original one. The classified focus points are accompanied by their probability of correctness. These values are taken from the error matrix of the classifier determined by analyzing the results on a representative sample of focus points. The probability values for focus points associated to a color region are added, renormalized and the hypothesis with maximum value is taken. To give a uniform interpretation for a region with associated focus point information and also Baufix element hypotheses generated by the isolated object recognizer, these results are integrated within the voting scheme. Therefore, the class system for the parts is condensed to the Baufix elements class system. Also the probabilities of correctness for the classified focus points are recalculated for the Baufix elements class system. The probability of correctness for an element hypothesis generated by the isolated object recognizer is assumed as one. By doing this the described plurality voting strategy is applicable to regions with associated focus point information and Baufix element hypotheses.

If not all the information sources agree in their hypothesis for a Baufix element, alternative interpretations are generated. Disagreements may occur on the one hand because of mistakes in the classification processes. On the other hand the object region of a small Baufix element, e.g., the head of a bolt, may be fully included within the region of a larger part, e.g., a bar. The *decision between the competing interpretations* is done by *knowledge based and probabilistic reasoning* on the alternative results for the whole bundle.

The assembly module generates from the hypotheses for Baufix elements and their associated
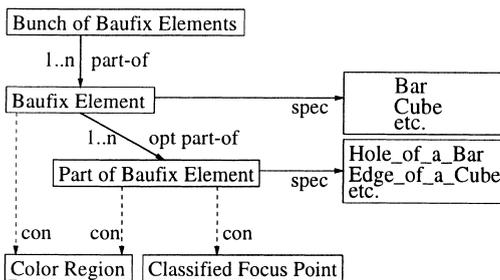

Fig. 5. Knowledge base for the semantic analysis within the central recognition unit.

regions an appropriate description for the assembly. If a possible structure is found for at least one alternative, those alternatives that do not form an assembly are rejected. The remaining alternatives are evaluated by taking into account the results of the voting scheme for the hypothesis of Baufix elements or parts of them. This evaluation is propagated through the net like a Bayesian propagation (Jensen, 1996) to find the best interpretation for the bunch of Baufix elements.

3. Results and outlook

The results for the integrated system for the example image is shown in Fig. 6. Comparing Fig. 6 with Figs. 2 and 3, it shows that the integrated system manages to overcome the error of the recognizer specialized for isolated elements in detecting the occluded cube by exploiting the focus point information. On the other hand, errors of the focus point classification, as the label *side_rhomb* for the light bolt and *tyre* for the bar are subdued by the integrated system for this example.

The module was tested on a sample of 63 isolated Baufix elements and 26 assemblies including 113 elements. The recognition results are evaluated by taking account of the correct element hypothesis and an appropriate associated color region (at least 80% overlap). The module specialized for isolated objects recognized 58 (92%) of the isolated and 66 (58%) of the assembled elements. From the 490 focus points marked altogether, 383 (78%) were classified correctly, 49 (10%) additionally as undefined points. The integrated system gave the correct result for 55 (87%) isolated and 80 (71%)
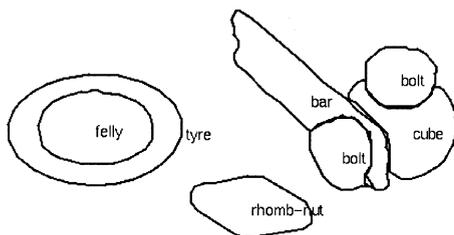

Fig. 6. Results of the integrated system.

assembled elements. These results emphasize that the recognition rate for assembled elements is increased by the integrated system. For isolated elements the results are decreased slightly, mainly because of errors of the classified focus points.

To further improve the system, we will integrate another working path dealing with contour information. This module provides additional information about object boundaries and structures within the object regions. Another important aspect will be exploiting all aspects of the top–down hypotheses from the assembly unit for the recognition task.

Discussion

Gelsema: You started out your presentation in a very general way, showing faces, houses and also these Baufix elements. I have the impression that the only domain to which you have applied your system is the domain of these Baufix elements. My question is the following: to what extent is the system general enough to also be applicable to other domains, in other words, how much effort would it take to have it applied to other domains?

Braun: This recognition scheme was implemented and tested on the Baufix scenario. The various modules are very different in their domain specificity. The module specialised for the recognition of isolated elements is especially domain specific, because there are different mechanisms that only work on Baufix elements. The modules for the detection of parts are more general. For example, the detection of focus points is a general-purpose module but it has to be parameterised, especially on the scale. The integration of element hypotheses is done by the semantic module and within the semantic module, for the moment, there is no special Baufix knowledge. There are just elements and parts of elements. I think that this is a general way to build up complex objects from their constituting elements. The main idea was to show that integration of different modules works well. For generalization, some of them have to be replaced by others, others can be parameterized. Combining such different modules will help in the recognition of complex objects.

Mulder: From your presentation it seems to me that you were somewhat surprised that the integration worked less well than expected. Is there an explanation for that? Usually you make assumptions about the world and I could imagine that the semantic model doesn't really fit to the complexity of Baufix elements.

Braun: One effect we have with the integration method is that it is slightly worse for the isolated elements, because the single module is really specialized for isolated elements without occlusions. But taking into account the simple semantic model and looking at the results for isolated elements and assemblies, we are not disappointed about the integration module.

Sethi: Could you tell me which are the illumination conditions in which you analyze your scene and how much your system is sensitive to a change of illumination?

Braun: There are different color region segmentation algorithms, and they are different in their dependence on illumination. The results I presented here are for fixed illumination.

Goode: Can you explain how your system would cope with variability in your elemental shapes? I understand that the Baufix modules have very well-defined shapes, but how would you cope with deformable shapes?

Braun: The well-defined shapes of the Baufix elements are used especially in the isolated element module, because the inputs are classification results and color regions. The results of the module which detects parts of elements are point based. The color region segmentation done here is general-purpose region segmentation. There are no restrictions to the shape. The combination is done on the basis of geometry, meaning that the points are integrated with the colors of the detected regions. Therefore, especially in this part, there are no explicitly formulated shape restrictions.

Gimel'farb: This approach depends heavily on the recognition of primitive elements. It seems to be a very complex problem to isolate such elements in the real examples (houses, faces), which

you have shown at the beginning. In the Baufix situation these elements and their colors are given by the manufacturers of the Baufix set. But, for instance, for a human face, how would one isolate the primitive elements of the face in such a way that the recognition of these elements will be simpler than the recognition of the face itself?

Braun: Face recognition suffers less from the problem of occlusion. Therefore, the recognition of elements may not be harder than the recognition of the face itself. But, for instance, in the example of the house, if the windows are occluded, I think it is perhaps necessary to decompose the objects into the primitive elements.

Gimel'farb: Have you ever experimented with more complex objects than those of the Baufix set?

Braun: No, the recognition scheme is implemented for the Baufix scenario.

Sagerer: Just one remark on the generality of this approach. The focus point approach, as well as the classification scheme is developed also for other problems. The description of assemblies is also developed for different problems. Also, the integration is developed for different problems. However, we have no second system which completely covers all aspects. We assume that it is a general scheme. So we really assume that this scheme is good enough for a number of pattern recognition problems in which you have complex objects in the sense that they are assembled like a house or a car, for which you can say a little bit about the parts.

References

Ballard, D., Brown, C., 1982. Computer Vision. Prentice-Hall, Englewood Cliffs, NJ.

Bauckhage, C., Kummert, F., Sagerer, G., 1998. Modeling and recognition of assembled objects. In: Proc. IEEE Conf. IECON98, Aachen, pp. 2051–2056.

Comaniciu, D., Meer, P., 1997. Robust analysis of feature space: Color image segmentation. In: Proc. IEEE Conf. CVPR, Puerto Rico, pp. 750–755.

Heidemann, G., 1998. Ein flexibel einsetzbares Objekterkennungssystem auf der Basis neuronaler Netze. Ph.D. Thesis. Universität Bielefeld.

Heidemann, G., Kummert, F., Ritter, H., Sagerer, G., 1996. A hybrid object recognition architecture. In: Proc. ICANN96, pp. 305–310.

Jensen, F.V., 1996. An Introduction to Bayesian Networks. Springer, London.

Parhami, B., 1994. Voting algorithm. IEEE Trans. Reliability 43 (4), 617–629.

Priese, L., Rehrmann, V., 1993. A fast hybrid color segmentation method. In: Pöppel, S., Handels, H. (Eds.), Mustererkennung 93, pp. 297–304.

Reisfeld, D., Wolfson, H., Yeshurun, Y., 1995. Context-free attentional operators: The generalized symmetry transform. Internat. J. Comp. Vision 14, 119–130.