TOWARDS AN INTEGRATED FRAMEWORK FOR CONTOUR-BASED GROUPING AND OBJECT RECOGNITION USING MARKOV RANDOM FIELDS

Daniel Schlüter, Sven Wachsmuth, Gerhard Sagerer

Bielefeld University
Faculty of Technology, Applied Computer Science
33594 Bielefeld, Germany
{dschluet,swachsmu,sagerer}@techfak.uni-bielefeld.de

Stefan Posch

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

ABSTRACT

We present an integrated approach for contour-based grouping and object recognition. Domain knowledge and domain-independent grouping laws are combined in a multi-layered Markov Random Field framework. It provides a basis for propagating top-down knowledge between different processing cues or input modalities. Additionally, the domain dependent MRF-layer can be used in order to evaluate the grouping process with regard to relevant contours for object recognition. Initial results show the approach to be adequate for complex scenes and partially occluded objects.

1. INTRODUCTION

Classical object recognition systems are decomposed into three tasks which are solved separately: image segmentation, feature extraction, and object classification. Moreover, they do not explicitly account for dependencies between features of the object. Recently, there have been some approaches [3, 1] which address these problems. Both use Markov Random Fields (MRFs) as a framework. Friedland and Rosenfeld describe the object recognition task as an optimization process of an energy function which combines low level and high level scores. Boykov and Huttenlocher capture pairwise dependencies between features of the object in the Gibbs distribution associated with a MRF. Both approaches define their energy functions on pixel level by using "multi-polar" representations or Hausdorff matching.

Our approach resembles some of these ideas, but uses a contour-based grouping hierarchy as an intermediate level of representation in order to reduce the degrees of freedom and the search space for higher level processing. Contour-based groups are first hypothesized using local information only, and subsequently judged within a more global context in a domain independent way employing ideas of the Gestalt psychologists for both steps. In contrast, the domain dependent MRF judges contour primitives top-down in the global context of an object hypothesis by applying geometrical constraints. Thereby, contour primitives are instantiated in the object model, and provide a basis for object class verification and 3D reconstruction.

2. CONTOUR-BASED PERCEPTUAL GROUPING

In this section we describe the basis of our approach, e.g. the contour segmentation and perceptual grouping.

The contour-based segmentation is realized with a standard Sobel operator applied to grey-scaled images, and with a subsequent thinning with non maximum suppression and hysteresis thresholding. The resulting edge elements are approximated with straight line segments and elliptical arcs with the method due to Leonardis [5].

The contour-based grouping organizes these contour segments within a hierarchy of grouping hypotheses using various Gestalt laws (Fig. 1). As a result fragmentation due to low contrast or occlusion can be resolved and more abstract image primitives are obtained.

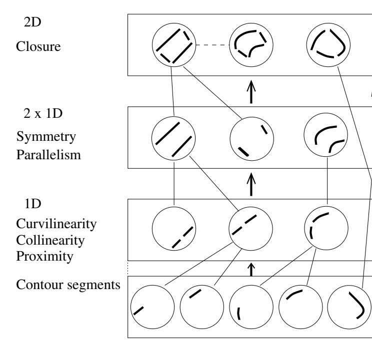


Fig. 1. Grouping hierarchy: Each group is depicted as a circle and thick arrows show the data flow of the grouping process. Some part of relations between groups are drawn with solid lines while competition is exemplified with a dashed line.

The hierarchy is divided into increasing levels of grouping complexity: In the lowest 1D level contour segments are combined to form collinear and curvilinear groups. These groups are again approximated with straight line segments or elliptical arcs and are as well as the initial contour segments denoted as *linear groups* subsequently. In addition, in the 1D level pairs of linear groups

In Proc. International Conference on Image Processing, volume II, pages 100-103, Vancouver, September 2000. IEEE.

may form a proximity grouping if their endpoints are spatially close. 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 labeled training set (details in [7]). In the highest level linear groups are organized into closed contours. To this end we construct an undirected graph from 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* modeling the underlying contours of a group, and *proximity edges* connecting two linear groups at their endpoints due to a proximity group. In this graph a closed contour corresponds to a simple cycle with an alternating sequence of contour and proximity edges which do not intersect.

The process of generating all hypotheses has a local character because it relies on the local evidence of the image data only. Therefore, all hypotheses are subsequently judged in a more global context using a domain independent MRF which will be described in the next section.

3. MARKOV RANDOM FIELD MODELING

Before we outline our use of Markov Random Fields for perceptual grouping and object modeling, we summerize the notation of MRFs used subsequently.

A given problem is represented with a MRF by defining a set of entities or *sites* $\mathcal{S} = \{s_1, \ldots, s_N\}$. Each site has to be assigned a *label* drawn from the label set $\mathcal{L} = \{l_1, \ldots, l_Q\}$ describing the state or interpretation of the site. To this end, a random variable X_i is associated with each site which accordingly takes values from the label set $\mathcal{L}$. The set of all *configurations* is denoted by $\Omega = \{\omega = (\omega_1, \ldots, \omega_N) | \omega_i \in \mathcal{L}\}$ where ω_i is the value of X_i. Interactions between sites result in a neighborhood system which can be expressed as an undirected graph with the sites as nodes. The maximum a posteriori estimate $\hat{P}(\omega|\mathbf{d})$ of the random field gives a globally consistent interpretation of the image data $\mathbf{d}$, which is calculated by minimizing the energy term $U(\omega|\mathbf{d})$ of the corresponding Gibbs distribution:[1]

$$U(\omega|\mathbf{d}) = \sum_{\mathbf{c}\in C} V_{\mathbf{c}}(\omega) - T \sum_{i=1,\ldots,N} \log P(d_i|\omega)$$

where C is the set of cliques and V_c are the associated clique potentials. In our approach the energy minimization is controled by “Highest Confidence First” (HCF) according to [2]. This method is deterministic, fast, independent of a specific initialization, and is shown to yield a local minimum which compares well to other minimization procedures. Other Approaches using MRFs for computer vision may be found in [6].

[1] We assume the independency of observations d_i at different sites s_i. This assumption of independency is violated in most practical applications, but common practice in computer vision, see e.g. [2].

3.1. Perceptual Grouping with a MRF

In our approach to judge hypotheses within the hierarchy, each grouping hypothesis is modeled as one site of the graph underlying the MRF. As a consequence, in contrast to most other approaches using MRFs, different sites may interpret a common subset of the image data. For example, a set of edge elements may be approximated by a line segment which is part of a collinear grouping hypothesis. This collinear grouping might in turn be part of a hypothesis for a closed contour. Thus, three sites of the MRF are related to a common set of pixels. Hence, the neighborhood system of the graph does not represent spatial neighborhood between grouping hypotheses, but rather models the dependencies between hypotheses with regard to a globally consistent interpretation of the image data. To this end, the graph is constructed with supporting and competing undirected edges between hypotheses: a grouping hypothesis is supported by all its parts, due to the fact, that a correct grouping hypothesis requires its parts as a prerequisite. On the other hand the hypothesis supports its parts, because it confirms them being a correct interpretation of the image data. In contrast, competing edges are defined between hypotheses, which represent contradicting interpretations of the image data, e.g. two collinear groupings compete, if they share common contour segments. Thus, supporting relations occur between the different levels of the hierarchy, while competing relations connect nodes within one level (Fig. 1).

The random variables associated with each site represent the significance of the hypothesis as being a correct interpretation of the image data, similar to the approach in [4]. The significance takes discrete values in the interval $[0\,;1]$, where large values encode high significance of a hypothesis. To complete the MRF, we have to define the a posteriori energy $U(\omega|\mathbf{d})$. Potentials of cliques of order two are used to model context-dependencies between sites, while data dependency is represented as single-node cliques. For example, supporting hypotheses which are both labeled as highly significant, obtain a negative clique potential. In contrast, if one of them is labeled with a high and the other one with a low significance a positive potential will be assigned, because this configuration is a contradicting interpretation for a supporting relation. Two hypotheses with low significance can neither be interpreted as contradicting nor consistent resulting in a neutral clique potential of zero. This qualitative definition is implemented with Gaussian smooth transitions between the different cases (details in [8]). Thus, minimizing the a posteriori energy of this MRF with HCF assigns a significance to each grouping hypothesis. Thresholding these significances yields *significant groups* as a globally consistent interpretation of the image data.

3.2. Object modeling

Each object of a given domain is derived from a 3D CAD-like model and represented by a set of 2D contour primitives for a given aspect. If we would know the correspondences of those model primitives to detected 2D image primitives, we can verify an object hypothesis and also can reconstruct their 3D positions [9]. In order to solve this correspondence problem, we use an MRF approach. Therefore, an *object MRF* (oMRF) for each object class is defined orthogonal to the grouping hierarchy. The sites are the contour based model primitives of a given object class (Fig. 2). The label set $\mathcal{L}$ is formed by significant linear groups resulting from minimizing the grouping MRF. The label set is extended by a special label ε in order to model object features which were not detected due to segmentation errors or occlusions. Thus, the correspondences between model primitives s_i and image primitives l_q is represented as an assignment of the l_q to site s_i. Geometric constraints between object primitives are explicitly modeled defining a neighborhood graph and associated clique-potentials. Currently, we model only cliques consisting of one or two nodes. The potential functions of those cliques are selected from a predefined set of parametric function types (a subset is shown in Fig. 3). The parameters can be estimated on a labeled test set – but are currently set by hand. The possible set of configurations is restricted to those which contain a given label $l_q \in \mathcal{L}$ only for one site. The optimization procedure for a given object class starts by dividing the label set $\mathcal{L}$ into subsets each consisting of the linear groups of a 2D closure $\mathcal{G}_i^{closure}$: $\mathcal{L}_i = \{l | l \in \mathcal{G}_i^{closure}, l \in \mathcal{L}\}$. Each label subset $\mathcal{L}_i$ is interpreted as an object hypothesis for all object classes which should be verified containing potentially additional groups. For verification we calculate the MRF configuration with the maximum a posteriori probability: Initially using only the label set $\mathcal{L}_i \cup \{\varepsilon\}$. After convergence, we add the remaining labels, e.g. image primitives not contained in $\mathcal{L}_i$, and continue minimization with the current configuration. Thereby, the HCF algorithm is directed to different local minima of the energy function representing different instances of an object class in the scene. The object hypothesis is accepted only if a sufficient number of model

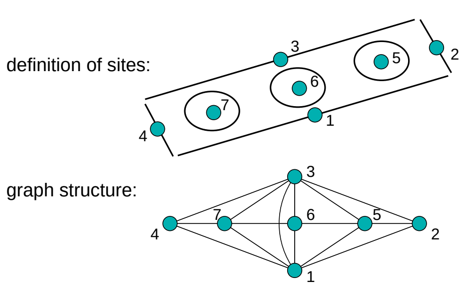
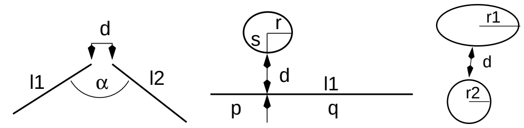


Fig. 2. object model of a three-holed-bar.



(a) adjacent lines (b) ellipse-line pairs (c) ellipse-ellipse pairs

$$V_c^{adj.lines}(\omega_1,\omega_2) = \max\{\sigma_{a,b}(d), \phi_{\mu,\sigma}(l2/l1), \zeta_{a,b,\mu}(\alpha)\}$$
$$V_c^{ell.lin.}(\omega_1,\omega_2) = \max\{\phi_{\mu,\sigma}(r/l1), \phi_{\mu,\sigma}(d/s), \phi_{\mu,\sigma}(p/q)\}$$
$$V_c^{ell.ell.}(\omega_1,\omega_2) = \max\{\phi_{\mu,\sigma}(d/(r1+r2)), \phi_{\mu,\sigma}(r1/r2)\}$$

Fig. 3. Subset of parametric function types for the definition of clique potentials. σ, ϕ, ζ are parametric weighting functions with inverted sigmoid, gaussian, or trapezoid shapes and a minimum at the optimal value.

primitives (sites s_i) have been assigned an image primitive from the label set ($w_i \in \mathcal{L}$).

4. RESULTS

Initial results have been calculated on real images showing complex assembled objects built from a wooden toy construction kit. The object recognition system should identify all three-, five-, and seven-holed bars that mainly characterize these assembled objects. In the first image (Fig. 5) two five-holed-bars have been tied up with two cubes. Both bars are partly occluded and, therefore, some relevant bar contours are broken or even not detected (Fig. 5b). Additionally, there is much noise and clutter in the initial contour segmentation caused by shadows. Nevertheless, the 2-D closures defining the silhouettes and holes of the two bars are reconstructed by exploiting the grouping hierarchy and both objects are correctly recognized. The algorithm has selected almost all relevant contours of these objects.

The second example (Fig. 4) shows the picture of an assembled toy-plane. On the one hand, it demonstrates the capabilities of the approach, on the other hand, some problems of the current implementation. The system correctly recognizes the three-holed bar defining the propeller of the toy-plane, the two seven-holed bars defining the wings, and the five-holed bar defining the tail-plane fin. The bar defining the body of the toy-plane is not recognized in Fig. 4c because no 2-D closure was found which could be used as a starting point for the object recognition procedure. In Fig. 4d we present the recognition results if this constraint is relaxed. Instead of starting the procedure for any 2-D closure it is started for any parallel lines which have been proved to be significant due to the MRF used for the grouping hierarchy. Now, all bars are correctly recognized including that defining the body of the toy-plane. This demon-

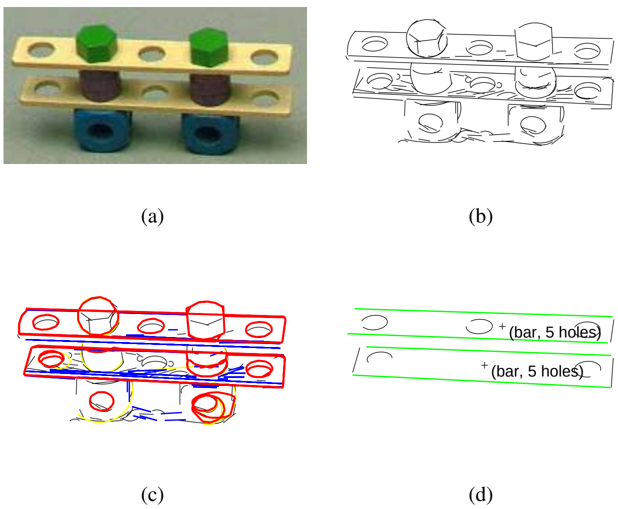


Fig. 4. Example of an assembled toy object: (a) original image, (b) initial contour primitives, (c) results of perceptual grouping[2], (d) selected linear groups with verified object hypotheses.

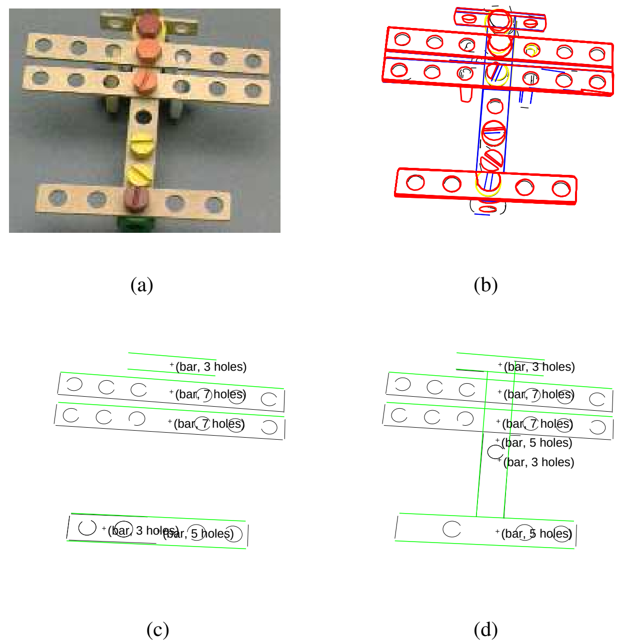


Fig. 5. Example of a plane: (a) original image, (b) results of perceptual grouping[2], (c) selected linear groups with verified object hypothesis based on significant 2D-closures, (d) selected linear groups with verified object hypotheses based on significant parallel lines.

strates the capability of our principal approach if the coupling between the grouping MRF and the object MRF is expanded to more than one level of the grouping hierarchy. Current drawbacks of using parallel lines as a starting point in general, are the expansion of the search space and a slightly higher rate of wrongly detected objects.

5. CONCLUSION AND OUTLOOK

In this paper we propose an integrated framework for contour based perceptual grouping and contour based object recognition using MRF modeling. The orthogonal definition of object class MRFs provides a basis to model more complex contour structures than feasible by a strictly domain-independent grouping hierarchy. Currently, the results of the grouping process are intensively used in the object verification process. As a next step, the information modeled in the object class MRFs will be used as top-down information to enhance judgement within the grouping MRF. Thereby, an object class hypothesis which may be provided by another input modality or cue can directly be translated to restrictions for the level of contour segmentation.

6. REFERENCES

[1] Yuri Boykov and Dan Huttenlocher. A new bayesian framework for object recognition. In *Proc. Computer Vision and Pattern Recognition*. IEEE, 1999.

[2] Paul B. Chou and Christopher M. Brown. The theory and practice of bayesian image labeling. *International Journal of Computer Vision*, 4:185–210, 1990.

[3] Noah S. Friedland and Azriel Rosenfeld. An integrated approach to 2d object recognition. *Pattern Recognition*, 30(3):525–535, 1997.

[4] S. Krishnamachari and R. Chellappa. Delineating buildings by grouping lines with MRF's. *Trans. on Image Processing*, 5:164–168, 1996.

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

[6] S.Z. Li. *Markov Random Field Modeling in Computer Vision*. Springer, 1995.

[7] A. Maßmann and S. Posch. Mask-oriented grouping operations in a contour-based approach. In *Proc. 2nd Asian Conference on Computer Vision*, volume 3, pages 58–61, Singapore, 1995. IEEE.

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

[9] G. Socher, T. Merz, and S. Posch. 3-D Reconstruction and Camera Calibration from Images with Known Objects. In D. Pycock, editor, *Proc. 6th British Machine Vision Conference*, pages 167–176, 1995.

[2] Closures marked red and parallels marked blue in online-version.