Int. Conference on Image Processing, Santa Barbara, 1997, Vol II, pp 207-210

Using Markov Random Fields for Contour-Based Grouping

Anke Maßmann, Stefan Posch, Gerhard Sagerer, Daniel Schlüter
AG Angewandte Informatik, Universität Bielefeld
P.O. Box 100131, 33501 Bielefeld, Germany
e-mail: posch@techfak.uni-bielefeld.de

Abstract

To overcome fragmentation of an initial contour-based segmentation and to organize contour segments into image primitives on a higher level of abstraction, regularities of the image data are exploited using ideas from the Gestalt psychology. First, groups are hypothesized within a hierarchy based on local evidence only, where the criteria are derived from a hand labelled training set. These hypotheses are subsequently judged in a global context using a Markov Random Field to derive a global interpretation. Examples of results for real data are given.

1 Introduction

The aim of this work is to exploit regular structure in a scene using ideas of the Gestalt psychologists, thus enhancing efficiency and robustness of subsequent processing steps of an image analysis system. Among others, this idea was proposed in [10]. We adopt a contour-based approach and initially segment the image resulting in contour segments approximated by straight line segments and elliptical arcs. These are used to define a hierarchy of grouping hypotheses with growing complexity using the Gestalt laws of proximity, good continuation, symmetry, and closure. Figure 1 shows the different levels of the hierarchy. The lowest level contains only one-dimensional primitives which are grouped according to collinearity, curvilinearity, and proximity. The medium level consists of symmetric and parallel grouping hypotheses, while the last level encloses hypotheses of closed contours. First, grouping hypotheses are constructed taking only local evidence into account. To judge these hypotheses a Markov Random Field is employed to include global constraints. MRFs have already been used for perceptual grouping in [9] at the signal level of grouping, according to the classification proposed in [8]. There, each site of the MRF corresponds to an individual pixel, while it is related to a symbolic image primitive (or feature) in our approach. In contrast to the MRF model for grouping of contour segments in [3], we model grouping hypotheses explicitly as nodes (or sites) of the MRF. Other approaches using MRFs for computer vision may be found in [5], related work dealing with grouping of contour segment in a hierarchical fashion in [7, 8] where a constraint satisfaction network and a Bayesian network is used respectively.

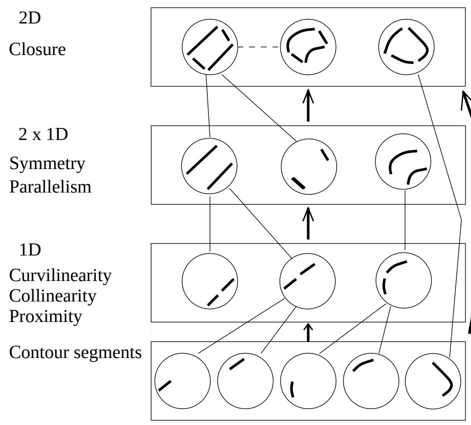


Figure 1: Hierarchy of grouping hypotheses: Each group is depicted as a circle, some examples of neighbourhood relations are given as thin solid lines (supporting) and thin dashed lines (competing). Thick arrows show the data flow of the generation of hypotheses according to the hierarchy. The dotted lines connecting the 1D level and initial segmentation indicate, that contour segments are considered as primitive groups.

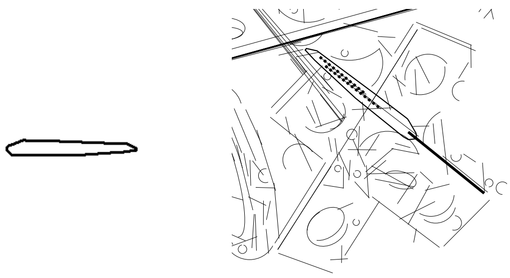
In the next section, an outline of the process generating grouping hypotheses within the hierarchy is given. Our model to use Markov Random Fields to represent the dependencies between hypotheses is described in section 3. Results are presented in section 4 and a summary is given in the concluding section.

2 Grouping Hypotheses on the Basis of Local Evidence

Grouping hypotheses are constructed bottom up with respect to the grouping hierarchy. The initial

In Proceedings International Conference on Image Processing, volume II, pages 207-210. IEEE, 1997.

segmentation is accomplished with a Sobel operator followed by a thinning with non maximum suppression and hysteresis threshold. The resulting edge elements are approximated with straight line segments and elliptical arcs with the method of Leonardis (see [4]). Pairs of contour segments may recursively yield 1D-grouping hypotheses, where each contour segment is considered as a primitive grouping of type collinearity or curvilinearity. Pairs of col- and curvilinear hypotheses may constitute a symmetry or parallelism which are represented in the 2x1D level of the hierarchy. To avoid combinatorial investigation of all pairs of segments and grouping hypotheses the search space is restricted to certain areas, which are called *Areas of Perceptual Attentiveness*. Information about shape and size of these areas is derived from a hand labelled training set of our domain for each grouping principle considered (see [6]). When applying an Area of Perceptual Attentiveness, it is first scaled and rotated according to the size and orientation of each individual segment or grouping hypothesis (see figure 2 for collinear groupings as an example). In addition, local criteria like orientation difference and distance are employed to restrict potential groupings. Again, the hand labelled training set is used to derive the necessary thresholds for these criteria.



(a) (b)

Figure 2: (a) Area of perceptual attentiveness for generating collinear grouping hypotheses; (b) Example of a rotated and scaled mask for a given line segment, which is display bold; two contour segments shown with thick dotted lines in addition fulfil orientation restrictions and are used to hypothesize two collinear groupings.

To hypothesize closed contours of the 2D level, a *proximity graph* is constructed from all contour segments, col- and curvilinear groupings as nodes and proximities between them as edges. From this graph closed contours are generated searching for simple cycles where the underlying contour segments do not intersect. Similarly, closed contours containing parallel contours are found by connecting the endpoints of parallel straight line segments or collinear groupings with a path of non-intersecting contour segments within the proximity graph.

3 Using Markov Random Fields for Global Judgement

The resulting grouping hypotheses are based only on local evidence and have to be judged in a more global context to yield a globally consistent interpretation. To this end we employ a Markov Random Field represented by an undirected graph.

Each grouping hypothesis corresponds to a node (or site) of this graph with an associated random variable. Therefore, 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 contour segment which is part of a collinear group hypothesis. This collinear grouping might in turn be part of a hypothesis for a closed contour. Thus, three nodes of the MRF are related to a common set of pixels. The same may be true for competing hypotheses, which are discussed below. The random variable of each node represents the significance of the hypothesis being a correct interpretation of the image data. It may take discrete values in the interval $[0, 1]$, where large values encode high significance of a hypothesis. In contrast to other approaches, the neighbourhood system of the graph does not represent spatial neighbourhood between grouping hypotheses, but rather models the dependencies between hypotheses with regard to a globally consistent interpretation of the image data. It is constructed with supporting and competing undirected edges: a 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 a correct interpretation of the image data. In contrast, competing edges are defined between hypotheses, which model contradicting interpretations of the image data: Tow collinear or two curvilinear groupings compete, if they share common contour segments. Two parallel groups or two closed contours compete with each other, if they again share a common part and in addition overlap a common region of the image. Thus, supporting relations occur mainly between the different levels of the hierarchy, while competing relations connect nodes within one level (see figure 1).

Markov Random Fields are characterized by a local property stating that the value of a random variable is only dependent on the value of its neighbours in the graph (see for example [2, 5]). A more practical approach to model this Markovian property is given by the Hammersley-Clifford theorem (see [1]), which states its equivalence to a Gibbs distribution. A Gibbs distribution is given as $P(\omega) = \frac{1}{Z} e^{-U(\omega)/T}$, where ω represents a given configuration of labels for each of the random variables associated to the nodes. Z is a normalizing constant and T represent the temperature, which controls the flatness of the distribution. The energy function $U(\omega)$ is defined as $U(\omega) = \sum_{\mathbf{c} \in \mathcal{C}} V_{\mathbf{c}}(\omega)$, where $\mathcal{C}$ is the set of cliques and $V_{\mathbf{c}}$ are the associated clique potentials. However, we are interested in the a posteriori probability $P(\omega|\mathbf{d})$ for the given image data $\mathbf{d}$ as observations. We assume conditional independence between the observations d_i at each site $s_i (i \in \{1, \ldots, N\})$, which is not exactly true but common practice in MRF approaches to computer vision (cf. [2]). Under this assumption and with Bayes's rule we can derive the a posteriori energy as:

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

Minimizing this a posteriori energy results in a maximum a posteriori estimate of the random field and gives a globally consistent interpretation of the image data. To complete the definition of the MRF the clique potentials $V_{\mathbf{c}}$ is developed in the rest of this section.

Potentials of singleton cliques are used to model the data dependency. To this end a data driven significance $data\_sig(s_i)$ is define for each hypotheses, which again maps to the interval $[0, 1]$. It defines a measure for the significance of a given hypothesis, which is based only on local evidence of the image data. Depending on the type of the grouping the approximation error, the gap between grouped segments and orientation differences are included. The clique potential measures the deviation of this local significance to the significance within the global context as assigned by the label ω_i of the current configuration:

$$V^{data}_{\{s_i\}}(\omega_i) = |\omega_i - data\_sig(s_i)|$$

The doubleton clique formulates the relations of support and competition between two given nodes. For supporting hypotheses the potential is designed as follows: two significant grouping hypotheses are consistent yielding a negative clique potential. A significant hypothesis and a hypothesis with low significance are contradicting and hence the assigned clique potential is positive. 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 a Gaussian

$$\psi(\omega_i, \omega_k) = \lambda \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}\left((\omega_i - \mu_i)^2 + (\omega_k - \mu_k)^2\right)},$$

where $\sigma = 0.2$ and $\lambda = 0.25$ yield smooth transitions between the different cases:

$$V_{\mathbf{c}}(\omega_i, \omega_k) = \begin{cases} -\psi(\omega_i, \omega_k) & : \ 0.5 < \omega_i, \omega_k \leq 1 \\ & \quad \text{with } \mu_i = 1,\ \mu_k = 1 \\ \psi(\omega_i, \omega_k) & : \ 0 \leq \omega_i \leq 0.5 < \omega_k \leq 1 \\ & \quad \text{with } \mu_i = 0,\ \mu_k = 1 \\ \psi(\omega_i, \omega_k) & : \ 0 \leq \omega_k \leq 0.5 < \omega_i \leq 1 \\ & \quad \text{with } \mu_i = 1,\ \mu_k = 0 \\ 0 & : \ 0 \leq \omega_i, \omega_k \leq 0.5 \end{cases}$$

In addition the supporting clique potential is normalized by the number of parts of a hypothesis. This is necessary, because for two competing hypotheses the one with more parts would receive more support just because of the larger number of supporting edges within the MRF graph. Therefore, the potential for supporting relations is defined as:

$$V^{support}_{\mathbf{c}}(\omega_i, \omega_k) = \frac{1}{NumParts(h_i)} V_{\mathbf{c}}(\omega_i, \omega_k)$$

The potential for competing relations is defined the other way round, since two competing hypotheses both being significant constitute an inconsistent interpretation. This is modelled by a positive contribution to the energy. In contrast, a significant hypothesis and a hypothesis with low significance are compatible and result in a negative potential.

$$V^{compete}_{\mathbf{c}}(\omega_i, \omega_j) = -V_{\mathbf{c}}(\omega_i, \omega_j)$$

Currently, cliques with more than two elements are not considered.

Given this definition of the clique potentials the a posteriori energy is summarized as:

$$U(\omega|\mathbf{d}) = \sum_{\{s_i, s_k\} \in \mathcal{C}_2^{support}} V^{support}_{\{s_i, s_k\}}(\omega_i, \omega_k) + \sum_{\{s_i, s_j\} \in \mathcal{C}_2^{compete}} V^{compete}_{\{s_i, s_j\}}(\omega_i, \omega_k) + \sum_{s_i} V^{data}_{\{s_i\}}(\omega_i)$$

The overall energy minimization is controlled by "Highest Confidence First" 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.

4 Results

Figure 3 (a) shows a typical scene of our domain, a wooden toy construction kit, and the initial segmentation (figure 3 (b)). The grouping hypotheses of type collinearity, curvilinearity, and closed contour are depicted in figure 3 (c). Proximity and parallel grouping hypotheses are not shown, since any representation results in a cluttered figure. The rather small number of collinear and curvilinear groups emphasise the quality of the initial segmentation. However, the fragmentation of the boundaries of wings and tail due to occlusion are overcome with appropriate collinear groups

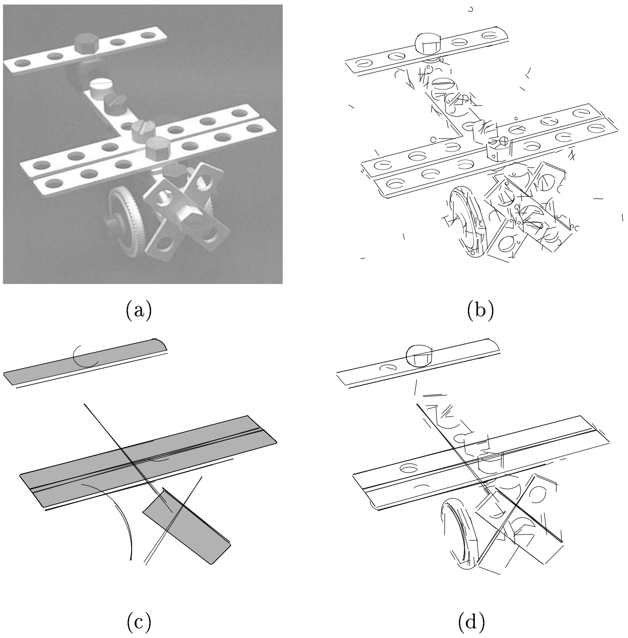


Figure 3: (a) Image of a wooden toy-plane; (b) initial segmentation; (c) generated hypotheses of collinear, curvilinear, and closed contours; (d) significant contour segment, collinear and curvilinear groupings

and closed contours. In addition to the two closed contours for each part of the wing, a third one for the whole wing unit was generated. For the occluded part of the air screw no closed contour is hypothesized, since part of the boundary at the right hand side is heavily fragmented due to low contrast. Therefore, only at the upper left hand side a collinear grouping hypothesis is able to bridge the gap due to occlusion. The judgement of these hypotheses in a more global context with the MRF results in a significance value for each grouping hypothesis. The contour segments, collinear and curvilinear groups with $\omega_i >= 0.6$ are shown in figure 3 (d). All closed contours beside the one for the complete wing unit are judged to be significant as well. Thus, the significant grouping hypotheses represent the salient parts of the aeroplane. Some detail is lost with respect to the holes in the wooden bars and to the body of the plane near the tail. The main reason for this are the contour segments, which do not belong to any grouping hypothesis and are therefore not considered by the MRF. For this example, the computation time on a DEC Alpha 500/400 is 15.8 seconds for the initial segmentation and 11.7 for generation and global judgement of the grouping hypotheses. The results for other images from this domain with varying complexity, zoom, and lightning condition show similar qualitative behaviour.

5 Conclusion

We introduced a hierarchical approach to perceptual grouping of contour segments. First, grouping hypotheses are generated locally employing Areas of Perceptual Attentiveness and thresholds derived from a hand labelled training set. These hypotheses are judge in a global context by a Markov Random Field modelling the dependencies between grouping hypotheses. The resulting significant contour groups represent the salient structure of the image and explicitly represent important closed contours as well. Thus, the grouping result is suitable as input for further processes like object recognition or 3D reconstruction. Further work regarding the design of the clique potentials, extension of the hierarchy, and coupling with a region-based approach is under way.

Acknowledgements

This work has been supported by the German Research Foundation (DFG) within SFB 360.

References

[1] Julian Besag. Spatial interaction and statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B*, 36:192–236, July 1974.

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

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

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

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

[6] 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.

[7] Rakesh Mohan and Ramakant Nevatia. Perceptual organization for scene segmentation and description. *Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, 14(6):616–635, 1992.

[8] Sudeep Sarkar and Kim L. Boyer. *Computing Perceptual Organization in Computer Vision.* World Scientific Publishing, 1994.

[9] Sabine Urago, Josiane Zerubia, and Marc Berthod. A markovian model for contour grouping. In *Proc. International Conference on Pattern Recognition*, volume A, pages 556–558, 1994.

[10] Andrew P. Witkin and Jay M. Tenenbaum. On the role of structure in vision. In Jacob Beck, Barabara Hope, and Azriel Rosenfeld, editors, *Human and Machine Vision*, pages 481–543. Academic Press, Orlando, 1983.