Image and Vision Computing 27 (2009) 854–863

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Figure-ground segmentation using factor graphs

Huiying Shen, James Coughlan *, Volodymyr Ivanchenko

Smith-Kettlewell Eye Research Institute, Rehabilitation Engineering Research Center, 2318 Fillmore St., San Francisco, CA 94115, USA

A R T I C L E I N F O

Article history:
Received 25 October 2007
Received in revised form 6 December 2008
Accepted 3 February 2009

Keywords:
Figure-ground segmentation
Belief propagation
Factor graphs
Text detection

A B S T R A C T

Foreground–background segmentation has recently been applied [S.X. Yu, J. Shi, Object-specific figure-ground segregation, Computer Vision and Pattern Recognition (CVPR), 2003; S. Kumar, M. Hebert, Man-made structure detection in natural images using a causal multiscale random field, Computer Vision and Pattern Recognition (CVPR), 2003] to the detection and segmentation of specific objects or structures of interest from the background as an efficient alternative to techniques such as deformable templates [A.L. Yuille, Deformable templates for face recognition. Journal of Cognitive Neuroscience 3 (1) (1991)]. We introduce a graphical model (i.e. Markov random field)-based formulation of structure-specific figure-ground segmentation based on simple geometric features extracted from an image, such as local configurations of linear features, that are characteristic of the desired figure structure. Our formulation is novel in that it is based on factor graphs, which are graphical models that encode interactions among arbitrary numbers of random variables. The ability of factor graphs to express interactions higher than pairwise order (the highest order encountered in most graphical models used in computer vision) is useful for modeling a variety of pattern recognition problems. In particular, we show how this property makes factor graphs a natural framework for performing grouping and segmentation, and demonstrate that the factor graph framework emerges naturally from a simple maximum entropy model of figure-ground segmentation.

We cast our approach in a learning framework, in which the contributions of multiple grouping cues are learned from training data, and apply our framework to the problem of finding printed text in natural scenes. Experimental results are described, including a performance analysis that demonstrates the feasibility of the approach.

1. Introduction

Originally proposed as a generic process for segmenting the foreground of a scene from the background, figure-ground (foreground–background) segmentation has recently been successfully applied [26,12] to the detection and segmentation of specific objects or structures (i.e. targets) of interest from the background. Standard techniques such as deformable templates [27] are poorly suited to finding some targets, such as printed text, stripe patterns, vegetation or buildings, particularly when the targets are regular (e.g. quasi-periodic) or texture-like structures with widely varying extent, shape and scale.

In these cases it seems more appropriate to group target features into a common foreground class (at least as an initial segmentation step to precede further processing), rather than directly attempt to find a detailed correspondence between a prototype and the target in the image, as is typically done with deformable template and shape matching techniques.

Our graphical model-based approach to figure-ground segmentation is an outgrowth of earlier work on figure-ground segmentation applied to finding crosswalks in traffic intersections [5] and builds on more recent work in [9,18,19]. The approach emphasizes the use of the *geometric* relationships of features extracted from an image as a means of grouping the target features into the foreground. In contrast with related MRF techniques [28] for classifying individual image patches into small numbers of categories, our approach seeks to make maximal use of geometric features extracted from images, rather than raw pixel information. Geometric information is generally more intuitive to understand than filter-based feature information, and it may also be more appropriate when lighting conditions are highly variable.

We formulate our approach in the general case of figure-ground segmentation and apply it to the problem of finding printed text in natural scenes. Experimental results are described, including a performance analysis that demonstrates the feasibility of the approach.

2. Grouping with factors

In this paper, we describe the use of factor graphs as a natural framework for grouping (i.e. segmenting) features in an image.

* Corresponding author. Tel.: +1 415 345 2146.
E-mail address: coughlan@ski.org (J. Coughlan).

0262-8856/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2009.02.006

Any grouping process analyzes relationships among features in deciding how to group them; these relationships are interactions among features that reflect their compatibility for inclusion into the same group. In much past work on grouping and segmentation, such as normalized cuts [22,26] and graphical model-based typical cuts [20] (all of which inspired our approach), these interactions are pairwise measures that measure the similarity (or affinity) of two features. Our approach is similar to that of [20] in that it constructs a Markov random field with one binary-valued (0 or 1) node variable for each feature to be grouped, and uses belief propagation to decide how to group the features (two features are assigned to the same group if the probability that they are jointly assigned to the same binary labels is above a threshold); however, their approach imposes a fundamental symmetry between the two possible states of each node variable, whereas in our approach the two possible states represent figure (state 1) or ground (state 0), which are not symmetric. The object-specific figure-ground technique in [26] imposes the same type of figure-ground asymmetry as our work does, but their work was applied to specific targets such as telephones and mugs and it is unclear how this technique would generalize to texture-like patterns with highly variable numbers of elements.

The above techniques all use *pairwise* measures that measure the similarity (or affinity) of two features. However, many clustering problems necessitate the use of higher-order interactions; for instance, the problem of grouping points on a 2D plane into lines requires an interaction defined on triplets of points, since every pair of points is trivially collinear. Some recent work [1] has investigated hypergraph partitioning techniques for handling these higher-order interactions.

Factor graphs [11] are graphical models designed to express interactions of any order (generalizing the pairwise interactions often used in graphical models, i.e. Markov random fields), and may be used for formulating simple and efficient grouping algorithms. We apply this formulation to the problem of object-specific figure-ground segmentation, which is how we cast the problem of text detection.

In the next section, we introduce factor graphs, and in subsequent sections we describe how a particular functional form of factor graph appropriate for figure-ground segmentation is suggested by a maximum entropy model of grouping.

2.1. Factor graphs

Factor graphs [11] are representations of probability distributions of many variables that can be expressed as a product of several factors, where each factor is an explicit interaction among a subset of these variables. (For a comprehensive treatment of factor graphs, see Chapter 8 of [2], which can be downloaded at http://research.microsoft.com/~cmbishop/PRML/.) Fig. 1 shows an example of a factor graph depicted in a graphical format. Each square node represents a factor, or interaction, among one or more variables, depicted by circles, and the topology of the factor graph indicates how the joint distribution of all variables factors.

The *arity* of a factor is the number of variables that interact in the factor. An arity-1 factor is called a *unitary* factor, and an arity-2 factor is sometimes referred to as a *pairwise interaction.* For the example shown in Fig. 1, factor h is arity-1, factor i is arity-2 and factors f and g are arity-3. In our figure-ground segmentation application, one factor arises from each constraint (measurement), and its arity equals the number of features (nodes) included in the measurement constraint.

The factor graph framework is an extremely general framework that is convenient for representing a large variety of probabilistic models. As we will see in a later section, inference on factor graphs is made tractable by approximate techniques such as belief propagation, which we use in our text-finding algorithm.

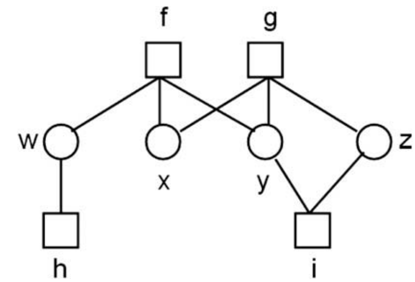

Fig. 1. Factor graph. This graph represents a distribution on four variables w,x,y,z (drawn as circles) using four factors f,g,h,i (drawn as squares). Edges connect factors with the variables they influence. The joint distribution represented by this factor graph is $P(w,x,y,z) = f(w,x,y)g(x,y,z)h(w)i(y,z)$.

2.2. Theoretical motivation: maximum entropy

In this section, we discuss the theoretical motivation for the functional form of the model we use for figure-ground segmentation, which is inspired from a simple maximum entropy probability model. As we will see, this functional form is naturally described using a factor graph.

The motivation for our maximum entropy model is that, given a collection of features to be segmented into figure or ground, evidence for how to assign (segment) features to figure or ground arises from considering *relationships among two or more features.* (Depending on the nature of the segmentation task, there may also be evidence for *individual* features belonging to figure or ground that is independent of other features.) The goal is to combine noisy evidence pertaining to groups of features in such a way that we can decide the likelihood that any *individual* feature should be assigned to figure or ground. Our approach is to construct a joint probability distribution of the assignments of all the features that is consistent with the evidence from groups of features. The joint distribution is chosen to be the least biased distribution that is consistent with the evidence; we enforce this "minimal-bias" criterion using a maximum entropy approach [17].

We now present the details of our maximum entropy model of figure-ground segmentation. We are given a collection of n features (i.e. nodes) in an image, each of which has an unknown assignment to figure or ground, denoted by x_i for $i = 1,2,\ldots,n$, where $x_i = 0$ represents assignment to the ground state and $x_i = 1$ represents assignment to the figure state. (Image data is also associated with each feature, such as its pixel coordinates.) Certain subsets of nodes are selected as *candidates to be grouped into the figure*, and we assume that for each subset (with at least two nodes) there is a way to estimate the probability that all nodes in the subset belong to the figure. Such a probability measure is based on the image data associated with the features; for instance, in the colinearity example in Section 2.4, the probability that a subset of three feature points belongs to the figure would be a monotonically increasing function of the degree of colinearity and proximity of the feature points.

We can regard the probability measure associated with each feature subset as a measurement that constrains the joint distribution of all nodes given the measurement data, which yields the posterior distribution $P(x_1,x_2,\ldots,x_n|data)$. (In this section, conditioning on "data" means that we are conditioning on some or all such measurement data; in later sections we will use more precise notation to refer to individual pieces of measurement data.) We can then show that the form of $P(x_1,x_2,\ldots,x_n|data)$ having the maximum entropy distribution consistent with these constraints will be a product of multiple factors, one factor corresponding to each constraint.