Pattern Recognition Letters 77 (2016) 42–49

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

# Shape-based object recognition via Evidence Accumulation Inference[☆]

Hui Wei[a,b,c,*], Qian Yu[a], Chengzhuan Yang[a]

[a] Laboratory of Cognitive Model and Algorithm, School of Computer Science, Fudan University, No. 825 Zhangheng Road, Shanghai 201203, PR China
[b] Shanghai Key Laboratory of Data Science, School of Computer Science, Fudan University, No. 825 Zhangheng Road, Shanghai 201203, PR China
[c] Shanghai Key Laboratory of Intelligent Information Processing, School of Computer Science, Fudan University, No. 220 Handan Road, Shanghai 200433, PR China

ARTICLE INFO

*Article history:*
Received 28 November 2014
Available online 8 April 2016

*Keywords:*
Shape-based object recognition
Evidence Accumulation Inference
Approximate inference
Greedy inference
Perceptual grouping
Attribute-centric object recognition

ABSTRACT

Shape-based object recognition is one of the most challenging problems in computer vision. Learning a structural representation using graphical models is a new trend in object recognition. This paper tries to apply graphical models to learn a shape representation and proposes a pipeline of shape-based object recognition. First, a Bayesian Network represents the shape knowledge of a type of object. Second, an Evidence Accumulation Inference with Bayesian Network is developed to search for the region of interest which is most likely to contain an object in an image. Finally, a spatial pyramid matching approach is used to verify the hypothesis to identify objects and to refine object locations. Our experiments corroborate that Evidence Accumulation Inference with Bayesian Network for object recognition is correct and show that the proposed pipeline achieves comparable results on well-known ETHZ shape classes and INRIA Horse dataset.



## 1. Introduction

Shape is a stable feature which is invariant to lighting conditions and variations in object color and texture. Current mainstream algorithms for object recognition mainly adopt appearance-based features, and ignored the shape features. People can easily distinguish between two objects' shapes without any extra information. However, it is difficult for an AI program designed for object recognition to achieve the human level. Shape-based recognition still remains one of the most challenging problems in computer vision. For the recent advance in contour detection proposed by Arbelaez et al. [1], shape-based object recognition in natural image is becoming more practical and attracts more attention in computer vision community.

The crucial assumption in cognitive psychology is that object recognition is mediated by recognition of the components of the object, which indicates that contour fragments are enough to be used to successfully recognize objects in an image. With this inspiration, we work to build a system of shape-based object recognition by utilizing the cues of contour fragments. The one of significant works of this paper is to learn a structural representation between pairs of contour fragments. A Bayesian Network is selected to build the structural representation of a type of shape, and is learned from a number of training examples.

One of major difficulties involves applying the structural representation. Directly utilizing the structural representation of shapes for object recognition will result in bad detections for the essence of the complication in natural images. We address this issue by proposing a pipeline with a multilayer architecture which is described in Fig. 3. The key step in this architecture is picking up region of interest (ROI) from an image by the proposed Evidence Accumulation Inference.

In summary, the main contributions of this paper include the following.

1. A structural representation of shapes, using Bayesian Network, is proposed which can be learned easily.
2. A novel way of Evidence Accumulation Inference is proposed to pick up region of interest which can reduce the number of False Positive and False Negative detections largely.
3. A pipeline of shape-based object recognition is built which achieves comparable results on well-known ETHZ shape classes and INRIA Horse dataset.

The remaining of the paper is structured as follows. The related works are reviewed in Section 2. In Section 3, encoding of contour segments is introduced. A shape knowledge representation is cre-

☆ This paper has been recommended for acceptance by Dr. N. Sladoje.

* Corresponding author at: Department of Computer Science, Fudan University, No. 825 Zhangheng Road, Shanghai 201203, China. Tel.: +86-021-51355515.
*E-mail address:* weihui@fudan.edu.cn (H. Wei).

http://dx.doi.org/10.1016/j.patrec.2016.03.022
0167-8655/

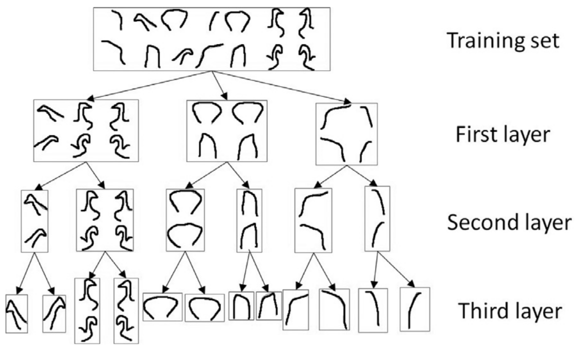


**Fig. 1.** Illustration of the process of hierarchical clustering of the training contour fragments into form codebook which consists of the cluster centers from each layer.

ated by Bayesian Network in Section 4. In Section 5, an Evidence Accumulation Inference method is used to search for region of interest, and the procedure of object detection is also described in detail. Experimental results are provided in Section 6. The final section gives the conclusion.

## 2. Related works

In recent years, a large number of shape-based object recognition methods have been also proposed and achieve the state-of-the-art performance on several well-known benchmarks. Ferrari et al. [2] organized the edges of images into Contour Segment Network and found paths through the network resembling the model outlines. Ferrari et al. [3] presented a family of scale invariant local shape features formed by short chains of connected contour segments, which were capable of encoding pure fragments of an object outlines. Ferrari et al. [4] proposed a novel technique for learning a shape model of an object class given images of example instances and combined Hough-style voting with a non-rigid point matching algorithm to detect the object in natural images. Ma and Latecki [5] proposed a scheme suitable for partial matching of edge fragments and localized objects on a weighted graph. Besides of literature mentioned above, Lu et al. [6] and Riemenschneider et al. [7] have also made comparable results on ETHZ shape classes.

The goal of the classical inference in Graphical Models is to solve the Marginal probabilities or Maximum a posteriori probabilities. The belief propagation algorithm [8] gives efficient and exact solutions to inference problems in tree-structured graphs. However, it is necessary to deal with graphs having loops. The junction tree algorithm [9] gives an exact inference on the arbitrary graph topologies. In most cases, it is not feasible to solve the inference problems with exact inference, and so it must develop some effective approximation methods. The representative approaches [10] to approximate inference include variational methods, sampling methods, loopy belief propagation, graph cut, linear programing methods, etc. This paper proposes a reverse inference problem and uses Evidence Accumulation Inference to solve it.

The philosophy of our method is related to the perceptual grouping approaches. Our method uses BN to represent a shape and develops a novel Evidence Accumulation Inference to collect contour fragments, whereas Havaldar et al. [11] proposed a method to detect contour fragments using perceptual organization criteria such as proximity, symmetry, parallelism, and closure. Selinger and Nelson [12] developed a hierarchy of perceptual grouping processes for 3d object recognition. Our method could also be related to the attribute-based object recognition approaches. Farhadi et al. [13] proposed an approach that went beyond naming and inferred attributes of objects. The attribute is similar to the proposed term, Evidence, in this paper. Farhadi et al. [14] employed BN to describe objects by the spatial arrangement of their attributes and the interactions between them. Wang and Ji [15] proposed a unified probabilistic model to incorporate attribute relationships. This paper proposes a shape representation to model the contour fragments instead of the object attributes with BN and a pipeline based on this shape representation is built for shape-based object recognition.

## 3. Annotation and encoding of contour fragments

Our system is performed on the contour images generated by gpb detector from [1]. The contour fragments are obtained by annotation. Each of them is described by three features and encoded with a codebook formed from the hierarchical clustering.

### 3.1. Annotation of contour fragments

As we know, a curve is better than a line to separate shapes. So each candidate of contour fragment must contain at least one critical point and may not be a line segment. There are two ways to obtain candidate fragments. The first is a way of auto-annotation which involves finding the critical points with the maxima of curvature along object contour and then identify curves between any pair of critical points as the candidates. In this paper, we detect critical points by Smoothed B-spline fitting. Auto-annotation is highly sensitive to the noise and it may generate a lot of false candidates, but is a nice choose for annotation on the large scale of benchmarks for the reason that it can reduce the work intensity of annotation. The second is a way of human annotation which involves manually labeling the contour fragments. Because strong supervision can improve the accuracy of annotation, it is a better choose to obtain the contour fragments generated by human annotator for small scale of benchmarks.

### 3.2. Descriptors of contour fragments

The curvature of the contour fragment is the best candidate for measurement which possesses invariance in translation, rotation, and scale transformation. The smoothed B-splines are used to fit points along a contour fragment. Then 64 points are sampled and their curvature values are calculated. In this way, we can represent each contour fragment as a vector $\mathbf{C_i} \in \mathbf{R^{64}}$. However, curvature features alone is not enough to describe a contour fragment. Two other features, Height–width ratio $\mathbf{R_i}\in\mathbf{R^2}$ and distance transformation $\mathbf{D_i}\in\mathbf{R^M}$, are here introduced to better depict the contour fragment.

### 3.3. Encoding of contour fragments

Encoding contour fragment features $\mathbf{F_i^T}=[\mathbf{C_i^T}, \mathbf{R_i^T}, \mathbf{D_i^T}]$ is to map feature vectors of contour fragments into a new space $\mathcal{B}$ spanned by a shape codebook $\mathbf{B}$; in this new space, each contour fragment is represented by shape code $w_i$. In order to form a codebook $\mathbf{B}$, the shapes in the training set are divided into contour fragments by annotator mentioned in Section 3.1 and then the aforementioned three descriptors are used to hierarchically cluster these contour fragments. Hierarchical clustering involved three layers as shown in Fig. 1. In the first layer, curvature features is used to cluster contour fragments into $\mathbf{M}$ classes. Then, for each class in the first layer, the depth-width ratio is used to further cluster contour fragments into $\mathbf{N}$ subclasses. Finally, each subclass in the second layer is clustered into $\mathbf{K}$ final classes by distance transformation. The clustering method used in each layer is k-means. The clustering centers in the three layer are $\mathbf{cen^1} = [\mathbf{cen^1_1}, \ldots, \mathbf{cen^1_{\mathcal{M}}}]$, $\mathbf{cen^2} = [\mathbf{cen^2_1}, \ldots, \mathbf{cen^2_{\mathcal{N}}} \times \mathcal{M}]$ and $\mathbf{cen^3} = [\mathbf{cen^3_1}, \ldots, \mathbf{cen^3_{\mathcal{K}\times\mathcal{N}\times\mathcal{M}}}]$,