Pattern Recognition Letters 000 (2016) 1–8

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

A probabilistic sparse skeleton based object detection[☆]

Burak Altinoklu[a], Ilkay Ulusoy[a], Sibel Tari[b,*]

[a] Department of Electrical and Electronics Engineering, Middle East Technical University, Ankara 06800, Turkey
[b] Department of Computer Engineering, Middle East Technical University, Ankara 06800, Turkey

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Markov random field
Generative shape model
Shape skeletons
Shape-based object detection

ABSTRACT

We present a Markov Random Field (MRF) based skeleton model for object shape and employ it in a probabilistic chamfer-matching framework for shape based object detection. Given an object category, shape hypotheses are generated from a set of sparse (coarse) skeletons guided by suitably defined unary and binary potentials at and between shape parts. The Markov framework assures that the generated samples properly reflect the observed or desired shape variability. As the model employs a sparsely sampled skeleton, the shape hypotheses are in the form of linear boundary segments; hence, matching can be performed using Directional Chamfer Matching. As the number of states that each MRF node can take is small, the matching process is efficient. Experiments with giraffe and swan categories of the ETHZ Dataset demonstrate that the method perform well in the case of articulated objects.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Deciding on whether any instance of a certain object category does or does not exist in a given image, and (if it does) estimating a bounding box containing the respective object instance is a task of practical importance. As such, this task, which is named as object detection, received attention from several researchers. There are two core interrelated computational processes: constructing a model of the object and searching the model in the image. A key decision to make is which cue to rely on for detecting an arbitrary instance of a certain object class. In this work, our focus is on detecting objects using only shape cues; more specifically, on shape based detection of articulated objects.

As we are particularly interested in articulated objects, it is important that the constructed shape model is flexible enough to generate possible variations of the shape due to articulated motion of the object and the model parameters expressing variations due to articulated motion are efficiently and effectively computable from a given set of example shapes (even when there are only a few). This motivates us to adopt a shape skeleton based model. Shape skeletons provide an abstraction of the object that expresses its articulating components and relations among them. Our skeletal representation is a set of sparsely sampled and ordered skeletal points with associated parameters to define the radii of the inscribed maximal balls and the boundary locations at which the respective maximal balls touch to. Hence, the object contour is represented by a finite number of boundary points producing a linear approximation (we later clarify the importance of linear approximation), which are located at the positions described by these parameters. The meaning of a point with the same index stays the same from one instance to another. The properties of an object class is captured with this representation. Therefore, it is possible to generate different shape templates of an object by varying few skeletal parameters.

While it is important that the model is flexible enough to generate possible variations of the object shape, it is also important that its conservative enough not to generate implausible shapes and that the generated shape hypotheses are easily incorporable to the image search process. This motivates us to consider a probabilistic framework using Markov Random Field (MRF) representation constructed with the help of our skeleton representation. Specifically, a node in the MRF denotes a part which is nothing but a skeletal branch comprised of a few sampled skeleton points. The edges denote pairwise relations between parts. Associated with each node i is a vector valued random variable X_i of which values determine a possible *linearized* instance of the shape of a part. Via suitably defined unary and binary potentials, $\phi(X_i)$ and $\phi\ (X_i, X_j)$, it is encouraged that generated shape hypotheses have certain regularity, *i.e.*, their variability correctly reflect the modeled (observed or forced) variability and the variations are consistent across parts (Fig. 1).

Once an MRF for the shape of a given object class is formed, the detection is formulated as MAP estimation problem via a joint

☆ This paper has been recommended for acceptance by Gabriella Sanniti di Baja.

* Corresponding author.
E-mail addresses: altinokluburak@gmail.com (B. Altinoklu), ilkay@metu.edu.tr (I. Ulusoy), stari@metu.edu.tr, sibeltari@gmail.com (S. Tari).

http://dx.doi.org/10.1016/j.patrec.2016.07.009
0167-8655/© 2016 Elsevier B.V. All rights reserved.

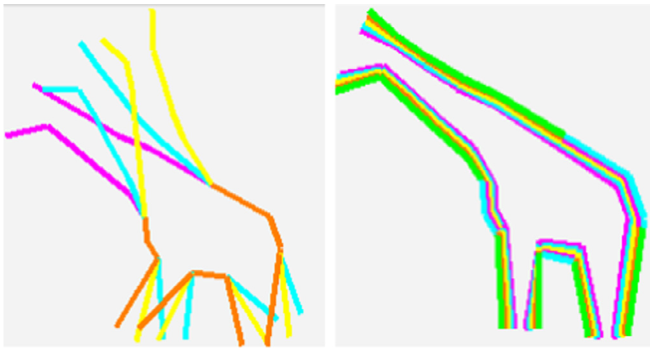

Fig. 1. Shape hypotheses generated by varying the two parameters separately.

distribution over the set of random variables $X = \{X_1, \ldots, X_{N_n}\}$, *i.e.*, finding the assignment of variables that maximizes joint posterior probability. The posterior probability contains a term based on image evidence; this term is Directional Chamfer Matching (DCM) cost between the linear shape hypothesis and the image edge map. In our case, the image edge maps are obtained via the Berkeley edge detector [16]. Using a DCM based cost on the edge map becomes possible because our shape hypotheses are in the form of a collection of lines. The maximization problem is solved using max-product belief propagation algorithm. Finding the assignment of variables that maximizes joint posterior probability is generally a complex problem; however, in the case of tree structured graphs, an efficient exact inference algorithm is available.

Related work

Most of the recent work in the literature use contour based representation of the shape of the object to be detected, e.g., [6,8,14,17,20,22,25]. When the shape of an object is modeled as a set of contour fragments, one may use *chamfer distance* function to match the set of contour fragments representing the shape of the object to the edge fragments detected from the image. The chamfer distance function, originally proposed in [4], is a smooth measure of the allignment of two contours. It is considerably robust to misalignment in position, scale and rotation. Hence, several shape based object detection methods employ chamfer distance for matching locally rigid contour fragments (modeling the shape of the object) to noisy edge maps extracted from images, e.g., [12,17,20].

An important distinction among methods is whether the object shape is represented via a single shape contour or via a model extracted from a collection of shapes. It is important that possible shape variations are taken into account. In the case of multiple instances for the shape of an object, a shape space can be constructed or shape parameters can be learned from the collection of instances. Even when only a single instance is available, it is important to model the object shape as flexible. This can be achieved by allowing some deformations. For example, in [17], a band around the shape is considered for tolerance.

One technical means of capturing shape variability, is to generate shape contours with the help of a skeleton model. Utilizing skeleton models brings a significant advantage if the object is an articulated one. As compared to contour based models methods, however, employing skeletons as shape models are only few. To the best of our knowledge, there are three methods. Bai et al. [3] consider a skeleton space as a collection of exemplars. Then they match boundary fragments corresponding to skeletal pieces to image edges using Oriented Chamfer Matching [20]. Adluru and Latecki [1] exploit the duality between the contour and skeleton. Trinh and Kimia [23] employ matched shock graphs to form a shape space. Shape hypotheses are checked in the image edge map using contour-partitioned chamfer cost.

Though there are only a few methods employing skeletons as shape models in object detection, the above mentioned alternatives commonly outperform purely contour based equivalents, especially for articulated objects. This is a motivation to further explore skeleton based object detection problem.

Contribution

Our contribution is to provide a simpler alternative to the most relevant competing method of Trinh and Kimia [23]. Even though Trinh and Kimia use dynamic programming for solution, their method overall is impractical. With their method, processing an image takes up to 20 min. There are three reasons for this: (1) there are more than 1 million states for each variable; (2) the partitioned Chamfer matching is slow; and (3) because the variability of the nodes are independent, implausible shape hypotheses are likely to be generated.

In our opinion, the weakest part of Trinh and Kimia [23] method is that its skeleton model is too complicated and the parameter sampling process is independent at each skeleton node. In comparison to it, our method offers significant computational advantages because of the following reasons. We use a simpler skeleton model with much less complex parameter search space. In our case, it suffices to use a directed chamfer matching which is faster. Hence, we can process images in seconds. Furthermore, our Markov framework makes it possible to generate more plausible shape hypotheses.

Organization

The rest of the paper is organized as follows. Our methods for model construction and detection are covered respectively in Sections 2 and 3. The dataset and the evaluation criteria are given in Section 4 and the experimental results are in Section 5. Finally, in Section 6, a summary is provided.

2. Model construction

In the model construction process, starting from at least one example shape, we arrive at the Markov Random Field representation of the object shape. The first step is to extract a reference skeleton based on one of the shapes. Then using the skeletons extracted from the remaining example shapes, we compute possible variations for the skeleton parameters. It is also possible that the range of the skeleton parameter variation can be set manually. We explain the details of our model construction process in three steps.

2.1. Reference skeleton

We extract the skeleton of the reference shape via Bai et al. [2] method, using its publicly available code. Because we use Directed Chamfer Matching later in the search step, the shape model needs to be approximated in the form of a collection of a sufficient number of line segments. The skeleton can be used for obtaining a linear approximation of the shape. First, we use the critical point detector of Kovesi [11] providing us with end and junction points, and then we trace the path between these critical points to obtain skeletal branches. Some of extracted critical points are close to each other. This creates a redundancy in the representation. Hence, we only keep a single end or junction point for those points of which pairwise distances are less than a threshold. Similar merging strategies are also used in [19] for obtaining stable skeletons across samples within a group. Manually, we fix a certain topology for each object class. For example, the skeleton structure consists

Download English Version:

https://daneshyari.com/en/article/4970204

Download Persian Version:

https://daneshyari.com/article/4970204

Daneshyari.com