PERGAMON

Pattern Recognition 36 (2003) 79–90

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Fast object recognition using dynamic programming from combination of salient line groups

Dong Joong Kang[a,*], Jong Eun Ha[b], In So Kweon[c]

[a]*Department of Robot System Engineering, Tongmyong University of Information Technology, Yongdang-dong 535, Nam-gu, Busan City, South Korea*

[b]*Samsung Corning Industry Corporation, Mechatronics Group, Kweonsun-gu, Suwon City, South Korea*

[c]*Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, 373-1, Gusong-dong, Yusong-gu, Daejun, South Korea*

Received 21 August 2001; accepted 28 January 2002

Abstract

This paper presents a new method of grouping and matching line segments to recognize objects. We propose a *dynamic programming-based formulation* extracting salient line patterns by defining a robust and stable geometric representation that is based on perceptual organizations. As the *endpoint proximity*, we detect several junctions from image lines. We then search for junction groups by using the *collinear constraint* between the junctions. Junction groups similar to the model are searched in the scene, based on a local comparison. A DP-based search algorithm reduces the time complexity for the search of the model lines in the scene. The system is able to find reasonable line groups in a short time. © 2002 Pattern Recognition Society. Published by Elsevier Science Ltd. All rights reserved.

Keywords: Feature matching; Dynamic programming; Perceptual grouping; Object recognition

1. Introduction

This paper describes an algorithm that robustly locates collections of salient line segments in an image. In computer vision and related applications, we often wish to find objects based on stored models from an image containing objects of interest [1–5]. To achieve this, a model-based object recognition system first extracts sets of features from the scene and the model, and then it looks for matches between members of the respective sets. The hypothesized matches are then verified and possibly extended to be useful in various applications. Verification can be accomplished by hypothesizing enough matches to constrain the geometrical transformation from a 3-D model to a 2-D image under perspective projection.

We first extract junctions formed by two lines in the input image, and then find an optimal relation between the extracted junctions, by comparing them with previously constructed model relations. The relation between the junctions is described by a collinear constraint and parallelism can be also imposed. Junction detection acts as a line filter to extract salient line groups in the input image and then the relations between the extracted groups are searched to form a more complex group in an energy minimization framework. The method is successfully applied to images with some deformation and broken lines. Since the system could define a *topological* relation that is invariant to viewpoint variations, it is possible to extract enough lines to guide 2-D or 3-D object recognition.

Conventionally, the DP-based algorithm as a search tool is an optimization technique for the problems, where not all variables are inter-related simultaneously [6–8]. In the case of an inhomogeneous problem such as object recognition, related contextual dependency for all the model features always exists [9]. Therefore, the DP optimization would not give the true minimum.

* Corresponding author. Tel.: +82-51-629-7224; fax: +82-51-629-7249.

E-mail addresses: djkang@tmic.tit.ac.kr (D.J. Kang), jeha2000@samsung.co.kr (J.E. Ha), iskweon@cais.kaist.ac.kr (I.S. Kweon).

PII: S0031-3203(02)00046-8

On the other hand, the DP method has an advantage in greatly reducing the time complexity for a candidate search, based on the local similarity. Silhouette or boundary-matching problems that satisfy the locality constraint can be solved by DP-based methods using local comparison of the shapes. In these approaches, both the model and matched scene have sequentially connected form of lines, ordered pixels, or chained points [10–12]. In some cases, there also exist many vision problems, in which the ordering or local neighborhood cannot be easily defined. For example, definition of a meaningful line connection in noisy lines is not easy, because the object boundary extraction for an outdoor scene is itself a formidable job as object segmentation.

In this paper, we do not assume known boundary lines or junctions but open any connection possibilities for arbitrary junction groups in DP-based search. That is, the given problem is a local comparison between pre-defined and sequentially linked model junctions and all possible scene lines in an energy-minimization framework.

Section 2 introduces the previous research about feature grouping in object recognition. Section 3 explains a quality measure to detect two line junctions in an input image. Section 4 describes a combination model to form local line groups as well as how junctions are linked with each other. Section 5 explains how related junctions are searched to form the salient line groups in DP-based search framework. Section 6 gives a criterion to test the collinearity between lines. Section 7 tests the robustness of the junction detection algorithm by counting the number of detected junctions as a function of the junction quality and whether a prominent junction from a single object is extracted under an experimentally decided quality threshold. Section 8 presents the results of experiments using synthetic and real images. Finally, Section 9 summarizes the results.

2. Previous research

Guiding object recognition by matching perceptual groups of features was suggested by Lowe [5]. In SCERPO, his approach is to match a few significant groupings from certain arrangements of lines found in images. Lowe has successfully incorporated grouping into an object recognition system. First, he groups together lines that are particularly likely to come from the same object. Then, SCERPO looks for groups of lines that have some property invariant with the camera viewpoint. For the purpose, he proposes three major line groups—proximity, parallelism, and collinearity.

Recent results in the field of object recognition including Jacobs, Grimson and Huttenlocher demonstrate the necessity of some type of grouping, or feature selection, to make the combinatorics of object recognition into a manageable level [13,14]. Grouping, as the non-accidental image features, overcomes the unfavorable combinatorics of recognition by removing the need to search the space of all the matches between image and model features. Grimson has shown that the combinatorics of the recognition process in cluttered environments using constrained search reduces the time complexity from an exponential to a low-order polynomial if we use an intermediate grouping process [14]. Only those image features considered likely to come from a single object could be included together in hypothetical matches, and these groups only need to be matched with compatible groups of model features. For example, in the case of constrained tree search, grouping may tell us which parts of the search tree to explore first, or may allow us to prune sections of the tree in advance.

This paper is related to Lowe's work using perceptual groupings. However, SCERPO grouping has a limitation: forming only small groups of lines limits the amount that may reduce search. Our work extends the small grouping to bigger perceptual groups including more complex shapes. Among Lowe's organization groups, the *proximity* consisting of two or more image lines is an important clue to start object recognition. When projected to the image plane, most man-made objects may have a polyhedral plane in which two or several sides give line junctions. First, we introduce a quality measure to detect meaningful line junctions denoting the proximity. The quality measure must be carefully defined so as not to skip salient junctions in the input image. Then, extracted salient junctions are combined to form more complex and important local line groups. The combination between junctions is guided by the *collinearity*, that is another perceptual groups (of Lowe's). Henikoff and Shapiro [15] effectively use an ordered set of three lines representing a line segment with junctions at both ends. In their work, the line triples or their relations as a local representative pattern broadly perform the object recognition and shape indexing. However, their system could not define the line triple when the common line sharing two junctions is broken by image noise or object occlusion. And the triple and bigger local groups are separately defined in low-level detection and discrete relaxation, respectively. The proposed system in this paper is able to form the line triple and bigger line groups in a consistent framework. Although the common line is broken, the combination of the two junctions can be compensated by the collinearity of the broken lines. We introduce the following: (1) define a robust and stable geometric representation that is based on the perceptual organizations (i.e., the representation as a primitive search node includes two or more perceptual grouping elements); and (2) introduce a consistent search framework combining the primitive geometric representations, based on the dynamic programming formulation.

3. Junction extraction

A junction is defined as any pair of line segments which intersect, and whose intersection point either lies on one of the line segments, or does not lie on either of the line