ELSEVIER

Expert Systems with Applications 31 (2006) 116–129

Expert Systems with Applications

www.elsevier.com/locate/eswa

Homomorphic graph matching of articulated objects by an integrated recognition scheme

Chin-Chung Huang [a,*], Innchyn Her [b]

[a] *Department of Mechanical Engineering, Yung Ta Institute of Technology and Commerce, 316, Chung Shan Road Linlo, Pingtung Hsien 909, Taiwan, ROC*
[b] *Department of Mechanical and Electro-Mechanical Engineering, National Sun Yat-Sen University, Taiwan, ROC*

Abstract

In the last few years, several attempts have been made to the study of object recognition under affine transformation, but all these studies have concentrated on graph isomorphism. There has not been any discussion of solving the graph homomorphism problem under affine transformation to date. Therefore, in this paper, an integrated approach, which combines the advantages of both the genetic algorithm and Hopfield neural network, is proposed for solving object recognition under this condition. The genetic algorithm is first used, to find the near-optimal solution including all the poses of the model in the scene. Then the Hopfield network is implemented and repeated to find each pose of the model. This method can solve occluded object recognition problems, and it can also obtain the homomorphic mapping indicating multiple occurrences of a model in the scene. The system is used to recognize articulated objects, and we do not need to know in advance that there is one in the scene. Kinematic properties like the position of the joint, relative displacement can be found for the articulated object. Through experiments, we can demonstrate that the proposed method is robust.

Keywords: Affine transformation; Hopfield network; Genetic algorithms; Object recognition; Graph homomorphism

1. Introduction

The geometric relation between 3-D objects and their views is a key component for various applications in computer vision. Perspective transformation is used to represent this relationship. It includes all possible viewpoint-related changes in images. However, it is difficult to recognize an object from a single image with perspective transformation. Fortunately, if the depth of the object perpendicular to the image plane is small compared with the viewing distance, perspective transformation can be regarded as a 'weak' perspective transformation, duly approximated by an affine transformation. Affine transformations contain typical geometric transformations, such as rotation, translation, scaling, and shear. They are used widely in the area of model-based object recognition.

Object recognition has been of considerable interest in the field of industrial automation. Many methods have been developed for the recognition of objects under affine transformation. They are Fourier descriptors (Arbter, Snyder, Burkhardt, & Hirzinger, 1990), moment invariants method (Huang and Cohen, 1996), B-spline invariant matching (Cohen, Huang, & Yang, 1995), geometric hashing (Lamdan, Schwartz, & Wolfson, 1990), and others. Among these methods, Fourier descriptors and moment invariants based methods are unable to handle the occlusion. Recently, Khalil, and Bayoumi (2001, 2002) used dyadic wavelet transform to derive affine invariant functions. Several invariant functions were presented by using different numbers of dyadic levels. Mokhtarian and Abbasi (2002) used the maxima of curvature scale space image to represent 2-D shapes. The shape was retrieved based on this shape descriptor, which is invariant under affine transformation. Although many of the recent methods do tolerate noise, they still cannot do things like recognizing multiple objects in the scene.

In the last two decade, there has been much research that incorporates neural networks for object recognition due to their robustness and the ability of parallel implementation. A good review on this subject can be found in (Egmont-Petersen, de Ridder, & Handles, 2002). In this paper, Egmont-Petersen et al. pointed out that it had hardly been considered explicitly regarding object recognition with occlusion by the neural classifier. Nasrabadi and Li (1991) first formulated this problem as a combinatorial optimization problem solvable by the Hopfield network. After that, some similar articles had been

* Corresponding author. Tel.: +886 8 7233733x228; fax: +886 8 7215649.
E-mail addresses: cchwang@mail.ytit.edu.tw (C.-C. Huang), her@mail.nsysu.edu.tw (I. Her).

0957-4174/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2005.09.005

presented to solve occluded graph matching problems (Kim, Yoon, & Sohn, 1996; Young, Scott, & Nasrabadi, 1997; Lee, Chen, Sun, & Tseng, 1997; Lin, Lee, Chen, & Sun, 1998). All of these approaches only recognized occluded objects under similarity transformations, and they were limited to the isomorphic mapping, which assumed that only one occurrence of the model is in the scene. Recently, Li and Lee (2001) presented a Hopfield method to solve graph matching problem under affine transformations. As the number of feature points is increased, their method will become unstable, and more spurious matching errors will be introduced. In a previous work of ours (Huang and Her, 2004), a primitive integrated scheme is presented to eliminate such spurious matching errors and to increase the robustness and efficiency of the recognition. However, Li and Lee's method and our early work were only suitable for isomorphic mapping. The homomorphic mapping, which permits multiple occurrences of the model in the scene, was considered by Suganthan, Teoh, and Mital (1995a,b, 1999). In their papers, the uniqueness constraint in the energy function was relaxed to obtain relational homomorphism, and consequently generated many spurious hypotheses. Moreover, their approaches were suitable for similarity transformations. It seems that there has never been any discussion in solving homomorphic graph matching problems under affine transformations.

The genetic algorithm (GA), on the other hand, is another effective tool for occluded object recognition. Recently, Khoo and Suganthan (2002, 2003) proposed a GA-based optimization procedure for the solution of structural pattern recognition problem. In their papers an attributed relational graph representation and a matching technique were used. They can also be used to solve homomorphic graph matching problem. However, all their papers can only work under Euclidean transformations.

As far as the applications of object recognition are concerned, most of the investigations to date have concentrated on the recognition of rigid, structure-like objects. However, most of the standard tools, as well as the industrial robots, possess internal kinematic degrees of freedom. Thus, in order to have a versatile object recognition system, one has to step further to acquire the ability of recognizing articulated rigid objects. An articulated object is an object composed of a set of rigid components, which are connected by joints and allow certain kinematic degrees of freedom. These joints can be, for example, prismatic joints that permit relative translation between components, or revolute joints that enable relative rotation of the components about the joint. The appearance and structure of an articulated object may change from view to view. Most of the object recognition systems do not have the capability of recognizing such objects. Only few articles have so far been presented to recognize articulated objects from intensity images (Beinglass and Wolfson, 1991; Hel-Or & Werman, 1994a,b; Bhanu and Ahn, 1998). However, if the articulated structure is unknown in advance, all these attempts will fail in such situation. More recently, Li and Lee (2002) presented an accumulative Hopfield matching to deal with the graph homomorphism problem and to recognize the articulated objects under similarity transformations. In their paper, the scene graph was divided into many subgraphs, and then the subgraph isomorphism is obtained. Unfortunately, all these methods would not work under affine transformations.

In this paper, we will further develop our integrated method that combines the GA and Hopfield network to recognize objects under affine transformations, and the objects can have multiple occurrences in a scene. We can also recognize articulated objects with a revolute joint or a prismatic joint. The objects need not be known as articulated in advance. The proposed method will find out the articulated objects and their kinematic parameters. Basically, the GA is first implemented to obtain a sub-optimal solution, which roughly includes the matching pairs of all poses. For even smoother integration, the GA is constructed as a specialized Hopfield-like model. A 2-D binary array is used to represent the genes. Accordingly, the initial state of the Hopfield network is set based on the results of the GA. The Hopfield network is executed several times if multiple occurrences of the model are detected in the scene. Thus, the final results of the Hopfield networks represent multiple poses of the model in the scene. Based on these multiple poses, the articulated object, if there is any, can be recognized. In this paper, we will test the performance of the proposed algorithm by recognizing some standard tools including pliers, scissors, vernier caliper, and adjustable wrench.

2. Discrete Hopfield nueral network

Our proposed method for solving homomorphic graph matching is to introduce an integrated approach that combines the use of the discrete Hopfield network (DHN) and the genetic algorithm (GA). Firstly, we want to discuss the use of the DHN for solving graph matching problem. If the model graph has M nodes and the scene graph has N nodes, an $M \times N$ 2-D binary neural network is to be constructed. The state of each neuron represents the measure of match between two feature points (nodes) of the model and the scene graph. The matching process can be characterized as to minimize the following energy function:

$$E = \frac{1}{2}\sum_i\sum_k\sum_{k\neq l} V_{ik}V_{il} + \frac{1}{2}\sum_i\sum_k\sum_{i\neq j} V_{ik}V_{jk} - \frac{1}{2}\sum_i\sum_k V_{ik} - \frac{1}{2} \times \sum_i\sum_k \sum_{\substack{(j\neq i)\\(j\neq i+1)\\(j\neq i-1)}} \sum_{\substack{(l\neq k)\\(l\neq k+1)\\(l\neq k-1)}} C_{ikjl}V_{ik}V_{jl}, \quad (1)$$

where V_{ik} is a binary output state of a neuron. If the ith feature point of the model object matches the kth feature point of the scene object, $V_{ik}=1$; otherwise, $V_{ik}=0$. In this paper, a neuron (i, k) is said to be active if $V_{ik}=1$. The first two terms of the energy function are used to enforce the uniqueness constraint