Obtaining Function Models using Active Vision Strategies

C. P. Lam
School of Engineering,
Murdoch University,
South Street,
Murdoch,
W. Australia 6150,
peng@eng.murdoch.edu.au

G. A. W. West and S. Venkatesh
School of Compting,
Curtin University of Technology,
GPO Box U1987,
Perth 6001,
Western Australia,
{geoff, svetha}@cs.curtin.edu.au

Keywords: Active Vision, Functional Representation, 2D and 3D Scene Analysis

Abstract

This paper addresses the problem of determining which 3D shape is present, and more importantly, the dimensions of the shape in a scene. This is performed in an active vision system because it reduces the complexity of the problem through the use of gaze stabilisation, choice of foveation point and selective processing by adaptively processing regions of interest. In our case only a small number of equations and parameters are needed for each shape. For example, a container has width and height. These are incorporated into functional descriptions of the shapes.

1 Introduction

A major objective of computer vision has been, and still continues to be, 3D object recognition. We live in a 3D world and though many useful applications of computer vision can avoid 3D object recognition, it is still widely regarded that if it is solved, then many applications are solvable. 3D object recognition has been pursued using a small number of basic techniques. Ignoring the use of shape from shading, stereo or motion and concentrating on a single camera and monochrome images, there have been three main methods: *recognition by components* [3], 2D appearance based systems, and 3D model based systems. In the 2D approach, a number of 2D views of an object are learnt and compared with the unknown image. Hopefully, the most appropriate view of the correct object will match best in some sense. Such systems use viewspheres [8] and characteristic views [2] to determine the optimum set of views of each object and use well known 2D matching and registration methods. In the 3D model based approach, a 3D model is manipulated to match the 2D projection of the model to the unknown image, usually by some form of gradient descent.

Many proposed 3D object recognition systems use a combination of the two techniques with 2D views serving as an index into most probably views of 3D objects. 3D matching is then used to refine the matches and firm up the hypotheses. Ultimately only one view of one object should be determined as a match.

All the above techniques suffer greatly from complexity when many objects have to be considered. The database consists of many accurate models and is complex if articulation, deformation and variations in shape are allowed.

An alternative approach is to rethink exactly what are the tasks of a 3D object recognition system. One is to consider function. The shape of a table is immaterial as long as it can satisfy the current goal, say, to support some object. This invokes ideas such as a flat surface, a certain height above the ground, stability and support. A cup is something that can contain a liquid and is of a suitable size with shape not being that important.

In this paper a technique is proposed that can recognise various functions. As such it deals with various shapes of objects. It can be compared with other techniques that can recognise static and rigid objects, or dynamic and deformable models, usually sequentially i.e. one at a time. The technique we propose can deal with parameterised models of various shapes defined by function. Shape varies as each function is parameterised e.g. containers have width and height. All models are treated simultaneously in parallel and the best chosen based upon minimising the error between hypothesised models and the sequence of images in an active vision system. This paper describes (i) how the models related to the objects in the scene are initially determined and (ii) the strategy used to improve the selected hypotheses. The strong point of this strategy is that the required parametric models and their associated parameters are determined in a dynamic manner. As each image is captured, the information obtained from it will be used to update the parameter values within the parametric models. Thus as