VARIOUS OBJECT RECOGNITION TECHNIQUES FOR COMPUTER VISION

LATHARANI T.R.1, M.Z. KURIAN2, CHIDANANDA MURTHY M.V.3

ABSTRACT

Object recognition in computer vision is the task of finding a given object in an image or video sequence. The task is monotonous in the dynamic world where the amount of visual information is enormous. This paper addresses some of the techniques available for object recognition for computer vision and their features. The different advantages and drawbacks found for each technique are highlighted. It also discusses the requirements of object recognition algorithms and the challenges ahead to be faced by them.

Keywords: Object Recognition, Learning module, Template matching, Clustering.

1. INTRODUCTION

Object recognition in computer vision is the task of finding a given object in an image or video sequence. It is a fundamental vision problem. Humans recognize a huge number of objects in images with little effort, even when the image of the objects may vary in different viewpoints, in many different sizes / scale or even when they are translated or rotated. Objects can even be recognized when they are partially obstructed from view. This task is still a challenge for computer vision systems. Recognition remains challenging in large part due to the significant variations exhibited by real-world images. In a dynamic world, information about the objects in the scene can become superseded even before it is ready to be used, if the recognition algorithm is not fast enough [7]. Partial occlusions, viewpoint changes, varying illumination, cluttered backgrounds, sensor noise and intra-category appearance variations all make it necessary to develop exceedingly robust models of categories.

Object recognition algorithms need to be "trained" using digital images. A large number of images must be gathered in order to correctly classify new objects. However, if we consider a large number of images each possibly containing an object of interest, it is impractical to search through all the images...
to identify a new object. The computational cost would grow with the number of images and therefore would not be close to real time. Research in object recognition is increasingly concerned with the ability to recognize generic classes of objects rather than just specific instances. Numerous methods for object recognition have been developed over the last decades [7].

2. LEARNING MODULE
The learning module’s[4] input is an image, its output is a label, for either the class of the object in the image (is it a cat?) or its individual identity (is it my friend’s face?). Learning module can be described as a binary classifier that gives an output of either “yes” or “no”. The learning module is trained with a set of examples, which are a set of input–output pairs, that is, images previously labeled.

In this setting, the distinction between identification and categorization is mostly semantic. In the case of categorization, the range of possible variations seems larger, because the system must generalize not only across different viewing conditions but also across different exemplars of the class (such as different types of dogs).

3. REQUIREMENTS OF OBJECT RECOGNITION ALGORITHMS
Reliable object recognition algorithms must fulfill the following important requirements:

Reliability: The algorithm should be robust with respect to noise and image variations, i.e. its performance should not degrade in case of slight variations of the object shape, poor image quality, changes in the lightning conditions etc.

Speed: As the applications may be designed to work online it implies certain demands on the speed of the image recognition process.

Automation: The process should be completely automated, i.e., it should work reliably without any additional user interaction.
4. CHALLENGES
The work is to visually recognize objects in real time for a large number of objects. Most existing object recognition systems suffer from poor scaling with the number of recognizable objects. Either the computational cost rises significantly or the recognition rate drops off as the number of objects increases. For this reason, most existing approaches have only been demonstrated on small data sets. Comparing and querying images require software with abilities similar to a human operator. The use of colors, textures, and interest points in the images are not enough, because two objects might be, for example, of different colors, dimensions and proportions. Query of images and video sequences must be done at the basic semantic level of identity of objects (e.g. tree, animal, house, human, etc.), while taking into account the context in which the object is found. The incapacity of designing such algorithms comes from the difficulty of dividing an image into its constitutive objects, from the difficulty of modeling shapes generically and from the difficulty of abstracting reflections, textures, shadows and distortions caused by the image capture process.

5. APPEARANCE BASED OBJECT RECOGNITION
Appearance-based object recognition systems are currently the most successful approach for dealing with 3D recognition of arbitrary objects in the presence of clutter and occlusion.

For appearance-based models, only the appearance is used, which is usually captured by different two-dimensional views of the object-of-interest. Based on the applied features these methods can be sub-divided into two main classes, i.e., local and global approaches [1].

A local feature is a property of an image (object) located on a single point or small region. It is a single piece of information describing a rather simple, but ideally distinctive property of the object’s projection to the camera (image of the object). Examples for local features of an object are, e.g., the color, (mean) gradient or (mean) gray value of a pixel or small region. For object recognition tasks the local feature should be invariant to illumination changes, noise, scale changes and changes in viewing direction, but, in general, this cannot be reached due to the simpleness of the features itself. Thus, several features of a single point or distinguished region in various forms are combined and a more complex description of the image usually referred to as descriptor is obtained. A distinguished region is a connected part of an image showing a significant and interesting image property. It is usually determined by the application of an region of interest detector to the image.

In contrast, global features try to cover the information content of the whole image or patch, i.e., all pixels are regarded. This varies from simple statistical measures (e.g., mean values or histograms of features) to more sophisticated dimensionality reduction techniques. Since the whole data is represented global
methods allow to reconstruct the original image and thus provide in contrast to local approaches, robustness to some extent. Contrary, due to the local representation local methods can cope with partly occluded objects considerably better.

![Fig. 2: Rectangles on the Left and Right Image are Examples of Possible Regions for Histogram Features. Stable Appearance in A, B and C on both Images Makes Corresponding Features to be Good Candidates for a Motorbike Classifier. On the Contrary, Regions D are Unlikely to Contribute for the Classification Due to the Large Variation in Appearance Here [2]](image)

This technique can be used with complex, curved shapes. It works extraordinarily good for recognition of a variety of 3-D shapes. But requires good whole-object segmentation, which limits their performance in the presence of clutter, occlusion, or background changes and also utilization simple conjunctions of low-level features causes crosstalk problems [1].

6. MODEL BASED OBJECT RECOGNITION

A model of an object undergoes some geometric transformation that maps the model into a sensor coordinate system [14]. Geometric algorithms draw results from computational geometry. The affine hashing method uses a redundant

![Fig. 3: Model of the Pliers and their Extracted Interest Points; (a), (b) Models of the Pliers and (c), (d) Extracted Interest Points](image)
representation of a set of points in order to locate that point set under an affine transformation, in the presence of extraneous data points.

An important characteristic of this method is division of the matching process into preprocessing and recognition stages. This significantly reduces the complexity of the algorithm compared with straightforward approach and enables execution of the preprocessing stage offline [8].

The model-based approach is well suited for an industrial environment since the objects processed by robot are usually known in advance and belong to a certain subset of the factory’s tools and products. This algorithm has been successfully implemented and tested on real-life data of industrial parts [8]. This technique provides a better running time, but sensitivity of the affine hashing approach to sensor errors is difficult to characterize.

7. Template Based Object Recognition

Template matching [9] is a technique in digital image processing for finding small parts of an image which match a template image. It can be used in manufacturing as a part of quality control [10], a way to navigate a mobile robot [11], or as a way to detect edges in images [12]. Template matching is a simple task of performing a normalized cross-correlation between a template image (object in training set) and a new image. For matching the template with the data image, different iterations of geometrical parameters (such as scale, rotation etc.) are applied and the required image is found. This method is normally implemented by first picking out a part of the search image to use as a template:
let the search image $S(x, y)$, where $(x, y)$ represent the coordinates of each pixel in the search image. Let the template $T(x_t, y_t)$, where $(x_t, y_t)$ represent the coordinates of each pixel in the template. Then simply move the center (or the origin) of the template $T(x_t, y_t)$ over each $(x, y)$ point in the search image and calculate the sum of products between the coefficients in $S(x, y)$ and $T(x_t, y_t)$ over the whole area spanned by the template. As all possible positions of the template with respect to the search image are considered, the position with the highest score is the best position. This method is sometimes referred to as ’Linear Spatial Filtering’ and the template is called a filter mask.

Above Figure shows an example with a real optical image. The first image shows a cup in a simple background. The second image shows the estimated position of the template with proper application of geometric parameters. It is seen here that the algorithm settled into this estimate after 42 iterations and that it is an accurate estimate [5]. Improvements can be made to the matching method by using more than one template, these other templates can have different scales and rotations. Template matching has various different applications and is used in such fields as face recognition and medical image processing. Systems have been developed and used in the past to count the number of faces that walk across part of a bridge within a certain amount of time. Other systems include automated calcified nodule detection within digital chest X-rays[13]. Template matching is a Hierarchical process, the required number of comparisons can be significantly reduced by clustering [15].

**Fig. 5: Template Matching [5]**

**Fig. 6: Sample Results [6] for Detection of a Face (a) and a Human Body (b) Each Image Shows the Globally Best Location for the Corresponding Object**
8. PART BASED OBJECT RECOGNITION
The basic idea is to represent an object by a collection of parts arranged in a deformable configuration. The appearance of each part is modeled separately, and the deformable configuration is represented by spring-like connections between pairs of parts. These models allow for qualitative descriptions of visual appearance, and are suitable for generic recognition problems.

9. REGION BASED OBJECT RECOGNITION
This algorithm transforms the input image into a directed graph which is constructed through several defined rules. The graph characteristics represent the global shape information of the object inside the input image, and are extracted during the graph construction. This strategy prevents the post processing traversing to the graph, with consequent improvement of computational time.

The algorithm was tested over a specific data base, and the experiments were conducted to show its performance in the light of two types of problems: object class recognition and similar image retrieval.

10. CONTOUR-BASED OBJECT RECOGNITION
The robot is equipped with an image database in which a variety of different objects is stored by single prototype images. Using a camera, the robot is expected to locate and identify by-passing objects, to pick them up and to put them to their final destination, thereby performing a task like sorting.

The object recognition strategy proposed here is a two-stage process. The first stage locates individual objects and describes them by polygonal shapes. It is able to cope with almost all kinds of shapes, including those with holes or strong concavities.

The second stage matches these shapes to the prototypes, detects the type of the objects, and computes their relative orientation. Alternative approaches to image segmentation combined with polygonization typically depend on initial edge detection processes and do not offer any strategies to effectively control the precision of the triangular approximation in a statistically well-motivated manner.

11. CONCLUSION AND OUTLOOK
The above attempts to recognize objects can be considered as typical in machine vision tasks. Often the requirements of an application are such that many assumptions can be made regarding the image under analysis. However, in unconstrained environments, recognition systems must deal with a variety of factors, and ultimately be capable of selecting one (or a few candidates) from amongst a large variety of possible shapes. No local representation of shape
can be completely invariant to loss of data, but what can be done is to ensure that the small change in input data results in a small change in representation. Continuous smooth variation in representation is necessary if we are to develop systems that cope with occlusions and are robust to the segmentation process. The representation must also have sufficient scope to deal with arbitrary shapes, even those which cannot be extracted as continuous boundaries. The issue of shape segmentation is of obvious importance, if we assume that object shapes come to us pre-segmented from a scene then subsequent recognition is almost trivial and the simplest approaches (measuring total length of lines etc.) will have some capabilities. If on the other hand we assume that we do not know which parts of an image are from a specific object the problem is almost impossible. Prior knowledge of the shape is clearly of some relevance to the recognition problem.

REFERENCES


[6] Pedro F. Felzenszwalb, “Pictorial Structures for Object Recognition”, Artificial Intelligence Lab, Massachusetts Institute of Technology, pff@ai.mit.edu Daniel P. Huttenlocher, Computer Science Department, Cornell University, dph@cs.cornell.edu.

[7] Arnau Ramisa1, Shrihari Vasudevan2, Davide Scharmaruzza2, Ramón López de Mántaras1 and Roland Siegwart2, “A Tale of Two Object Recognition Methods for Mobile Robots”, 1 Artificial Intelligence Research Institute (IIIA-CSIC), Campus UAB, 08193 Bellaterra, Spain, 2 Autonomous Systems Lab, ETH Zurich, Switzerland.


