CHAPTER-1

Object Recognition For Robot Vision

1.1 Introduction

The mapping of real world objects to the image plane including the geometric and the radiometric parts of image formation is basically well understood. The image data and prior knowledge of the considered application are the major sources of information for various vision issues. Common examples are image restoration, filtering, segmentation, reconstruction, modeling, detection, and recognition or pose estimation algorithms. The hardest problems in computer vision are related to object recognition. Up to now, there have been no general algorithms that allow the automatic learning of arbitrary 3-D objects and their recognition and localization in complex scenes. State-of-the-art approaches dealing with high-level vision tasks are essentially dominated by model-based object recognition methods. In the past, numerous applications have led to various types of representations for object models that allow the implementation of excellent recognition systems.

Many image processing and computer vision algorithms are characterized by prevailing adhoc solutions. Most techniques apply intuitive ideas specified for the given application and neglect the exploration of precisely defined mathematical models. There exists no unified theoretical formalization that provides the framework for the analytical analysis of designed complete systems. For the most part, it is left to empirical studies to justify the usage of the chosen representation scheme. We also cannot introduce general models that fit all requirements of conceivable applications, but we do present some probabilistic modeling schemes and their basic features, which have been shown to be proper for a wide range of vision problems.
1.2: Object Recognition System

An object recognition system finds objects in the real world from an image of the world using object models, which are known as a priori. Human performs object recognition effortlessly and instantaneously. Algorithmic description of this task for implementation on the machine has been very difficult. The problem of object recognition can be defined as a labeling problem based on models of known objects. The problem of object recognition can be closely related to segmentation problem, without at least partial recognition of objects, segmentation cannot be done and without segmentation object recognition is not possible. The basic components of object recognition systems are

- Model database (Model base)
- Feature detector
- Hypothesizer
- Hypothesis verifier

The component as connected to each other are as per shown in figure 1.1

![Diagram](image)

Figure 1.1: Basic Components of Object recognition System.

The Model database contains all the models known to the system. The information in the model database depends on the approach used for the recognition. It can vary from a quantitative or functional description to precise geometric surface information. In many cases
the models of objects are abstract features vectors. A feature is some attribute of the object that is considered important in describing and recognizing the object in relation to other objects. Size color and shape are some commonly used features.

The feature detector applies operators to images and identifies locations of features that help in forming object hypothesis. The feature used by the system depends on the types of objects to be recognized and the organisation of the model database. Using detected features in the image the hypotheses assigns likelihood to objects present in the scene.

The Modelbase is organized using some type of indexing scheme to facilitate elimination of unlikely object candidates from possible considerations. The verifier then uses object models to verify the hypothesis and refines the likelihood of objects. The system then selects the object with the highest likelihood, based on all the evidence as the correct object.

1.3 Complexity of Object Recognition

An image of scenes depends on illumination, Camera parameters and camera location. The object has to be recognized from images of a scene containing multiple entities. The complexity involved in Object recognition can be because of factors like:

a. **Scene Constancy:** The scene complexity will depend on whether the images are acquired in similar conditions (e.g. illumination, background, camera parameters and viewpoint) as the models. Under different scene conditions, the performance of different feature detectors will be significantly different.
b. **Image model spaces**: Many times 3-dimensional objects have to be represented using 2-dimensional features models. Such models are can be represented using 2-dimensional characteristics e.g. if the models are 3-D and perspective effects are to be considered the problem of feature detection becomes more complex.

c. **Number of models in the model database**: If the numbers of objects are very small then hypothesis formation stage can be ignored and sequential exhaustive matching may be used for object recognition. Hypothesis formation becomes important for a large number of objects and amount of efforts spent in selecting appropriate features becomes more complex.

d. **Number of objects in an image and possibility of occlusion**: If there is only one object in an image it may be completely visible. With an increase in the number of objects in the image, the probability of occlusion increases. Occlusion is a serious problem in many basic image computations. Occlusion results in the absence of expected features and the generations of unexpected features.

If the images of the object can be obtained from arbitrary viewpoints, then an object may appear very different in its multiple views. For object recognition using 3-D models, the perspective effects and viewpoints of the image have to be considered. Even if the models are 3-D and the images contain only 2-D information affects object recognition approaches.
A structure of typical pattern recognition system is as shown in the Figure 1.2

1.3.1 Procedure for P.R. System Engineering

<table>
<thead>
<tr>
<th>STEP</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Study the classes of pattern under consideration to develop possible characterisation. This includes assessment of pattern structure and probabilistic characterisation as well as exploration of possible within class and inter class similarity/Dissimilarity measures. In addition possible pattern deformations or invariant properties and characterisation of noise sources should also be considered at this point.</td>
</tr>
<tr>
<td>2</td>
<td>Determine the availability of features / measurement data.</td>
</tr>
<tr>
<td>3</td>
<td>Consider constraints on desired system performance and computational resources.</td>
</tr>
<tr>
<td>4</td>
<td>Consider the availability of training data.</td>
</tr>
<tr>
<td>5</td>
<td>Consider the availability of suitable and known PR techniques and overall PR system structure</td>
</tr>
<tr>
<td>6</td>
<td>Develop PR system simulation. This may involve choosing Models, Grammar or network structure.</td>
</tr>
<tr>
<td>7</td>
<td>Train the system</td>
</tr>
<tr>
<td>8</td>
<td>Simulate system Performance.</td>
</tr>
<tr>
<td>9</td>
<td>Iterate among the above steps until desired performance is achieved.</td>
</tr>
</tbody>
</table>
1.4 Different Pattern recognition approaches

Pattern recognition problem can be classified on the basis of the technique used for pattern classification and can be categorised as

1. Statistical Pattern Recognition
2. Syntactic Pattern Recognition

Out of these techniques statistical pattern classification technique is being used for classification where large features are involved in making decision. The general structure of a statistical Pattern recognition system is as shown in the figure 1.3. Statistical method of classification is also called as decision theoretical technique.

![Figure 1.3: Statistical Pattern recognition System.](image_url)

Selection of the feature from the feature vector is difficult, but should be chosen in such a way that they should be less sensitive to variations in the pattern. Selection is based upon experiments with a training set of known patterns. A feature vector containing N features maps each pattern as a point in an N dimensional feature space. The decision about whether the given pattern belongs a class or not is made by using bayes decision rule, which can be stated as.

\[ P(\omega_j|x) = \frac{p(x|\omega_j)p(\omega_j)}{p(x)} \quad \text{Where} \quad p(x) = \sum_{j=1}^{n} p(x|\omega_j)p(\omega_j) \quad ......(1.1) \]

and \( p(\omega_j) \) is priory probability. \( P(x|\omega_j) \) is conditional density function and \( p(\omega_j|x) \) is post priory probability. Based on this rule pattern classifier can be design. There are two approaches for
implementation of classifier they are

1. Parametric classification
2. Non-parametric classification.

In parametric techniques we use the samples to estimate the unknown probability and probability density function and use the resulting estimate, as they are the true values. We can parameterise the conditional density and reduce the dimensionality problem. We also make certain assumptions about the form of density function and based on that assumption the classifier can be built, but in nonparametric method neither the density function is known nor it is possible to parameterise the given problem. Hence in this technique first approach is to estimate the density function from the available set of known samples and using this function classifier can be design. The different techniques, which can be used for density estimation, are

1. Parzen window.
2. Kn-Nearest-neighbor Estimation
3. Approximation by series expansion.

Since the density function has to be estimated from the known samples available we can use neural network techniques and train the classifier so that the decision about to which class a input pattern belongs to can be made. Neural nets can be used to implement the classifier. Their attraction lies in their ability to partition the feature space using non-linear boundaries for classes. These boundaries are obtained by using training of the net.

1.5 Robot Vision

Robot vision may be defined as the process of extracting, Characterising and interpreting the information from images. CCD (Charge Coupled Device) cameras are commonly used to capture the image. Most of the CCD camera discretise the scene into 512 X 512
format which can be stored in the Computer. A typical robot assisted system is as shown in the figure 1.4. This system comprises of a 5-axis robot driven and powered by its own controller. A CCD camera views the scene and this scene is stored in computer memory which can be further processes as per the requirement of the user by using a image processing card. The image is represented into 512 X 512 picture elements and each pixel element is called as pixel. Each pixel takes a value between 0 to 255 depending upon the intensity of illumination of that element. This value is also called as grey level. This image then is stored in 512 X 512-memory arrays and can be used for further processing. The raw image stored in the computer needs some processing so that some useful data can be extracted out of this image this is called pre-processing. For pre-processing normally simple masks (templates), Windows or filter can be used. Some pre-processing techniques, which are normally employed on raw image, is Image Smoothing, Median Filtering, Smoothing of Binary images and thresholding.

Machine Vision (Also known as Computer Vision, robot vision or artificial vision) is an important sensor technology with potential applications in many industrial operations. Many of the current applications of machine vision are in inspection. But as there are factors like

1. Constantly reduction in cost of computational hardware.
2. Increase in speed of processing.
3. Constant development of new and better algorithms being developed

It is anticipated that vision technology will play an increasingly significant role in the future of robotics.
For an efficient vision system two important criteria must be met so that the machine vision systems can be employed in functions like.

1. Selection and isolation of machine parts
2. Identification of Parts.
4. Complex inspection for closed dimensional tolerances.
5. Improved recognition and part recognition
6. Vigilance and security systems.

These criteria are
1. The system so designed should be relatively economical
2. The system should have relatively rapid response.
Predictions are that the field of computer vision will be one of the fastest growing commercial areas in the twenty first centuries. Computer vision is a complex and multidisciplinary filed and is still in its early stages of development. Advances in the vision technology and related discipline are expected within the next decade which will permit applications not only in manufacturing but also in photo interpretation, Warehousing, robotics operations in hazardous environments, autonomous navigation, cartography and medical image analysis.

Machine vision is concerned with the sensing of data and its interpretation by a computer. The typical vision system consists of,

1. Camera and digitising hardware.
2. A digital computer.
3. Hardware and software to interface them.

Operations of the vision system consist of three functions.

1. Sensing and digitising image data
2. Image processing and analysis.
3. Applications.

Sensing and digitising function involves the input of vision data by means of a camera focused on the scene of interest. Special lighting techniques are frequently used to obtain an image of sufficient contrast for further processing. Image viewed by the camera is typically digitised and stored in computer memory. Digital image is called a frame of vision data and is frequently captured by the hardware device called as frame grabber card. The scene captured by the card is stored as elements of a matrix and is called as pixels. A sampling process performed on each image frame determines the
number of pixels. A single pixel is the projection of a small portion of
the scene, which reduces that portion to a single value. The value is
measure of light intensity for that element of the scene. The digitised
image matrix for each frame is stored and then subjected to image
processing and analysis functions for data reductions and
interpretation of the image. These steps are required in order to
permit the real time applications of vision analysis required in
robotics applications. Image frame will be threshold to produce the
binary image and then various feature measurements will further
reduce the data representation of the image. This data reduction can
change the representation of the frame from several hundred
thousand bytes of raw image data to several hundred bytes of feature
value data. Resultant feature data can be analysed in the available
time for the action by the robot system.

To accomplish image processing and analysis the vision system
frequently must be trained. In training information is obtained on
prototype objects and stored as computer models. The information
gathered during training consists of features such as area of the
object, its perimeter length, Major and minor diameters and many
such features. Feature values computed on unknown objects viewed
by the camera are compared with the computer models to determine
if a match has occurred.

Three dimensional vision systems may require special lighting
techniques and more sophisticated image processing algorithms to
analyse the image. Many three dimensional recognition systems
require two cameras in order to achieve a stereoscopic view of the
scene. The image so obtained needs to be processed so that some
useful information can be obtained for further processing. The
various processing activities, which can be operated on the image,
are.
Image data reduction: The main objective of this processing is to reduce the volume of the data. Following two schemes have found usage for data reduction.

1. Digital conversion
2. Windowing.

Segmentation: Objective of this processing is to group areas of an image having similar characteristics or features into distinct entries representing parts of the image, like Boundaries (edges) or regions (areas). The various techniques, which can be employed to achieve this, are.

1. Thresholding
2. Region growing
3. Edge detection.

1.6 Pattern recognition for computer vision

The basic goal of signal processing in computer vision is the extraction of "suitable features" for subsequent processing to recognize and classify objects. But what is a suitable feature? This is still less well defined than in other applications of signal processing. As signals processed in computer vision come from dynamical 3-D scenes, important features also include motion and various techniques to infer the depth in scenes including stereo, shape from shading and photometric stereo, and depth from focus.

There is little doubt that non-linear techniques are crucial for feature extraction in computer vision. However, compared to linear filter techniques, these techniques are still in their infancy. There is also no single non-linear technique but there are a host of such
techniques often specifically adapted to a certain purpose.

In principle, *pattern classification* is nothing complex. Take some appropriate features and partition the feature space into classes. Why is it then so difficult for a computer vision system to recognize objects? The basic trouble is related to the fact that the dimensionality of the input space is so large. In principle, it would be possible to use the image itself as the input for a classification task, but no real-world classification technique be it statistical, neuronal, or fuzzy would be able to handle such high dimensional feature spaces. Therefore, the need arises to extract features and to use them for classification.

Unfortunately, techniques for feature selection have very often been neglected in computer vision. They have not been developed to the same degree of sophistication as classification, where it is meanwhile well understood that the different techniques, especially statistical and neural techniques, can be considered under a unified view. Morphological operators can help us to extract some features from the image because they manipulate the shape of objects in images. Fuzzy image processing contributes a tool to handle vague data and information. Object recognition can be performed only if it is possible to represent the knowledge in an appropriate way. In simple cases the knowledge can just rest in simple models.

1.7 Why probabilistic models.

An immediate question is why we should prefer a probabilistic setup to any other recognition algorithms. In fact, both from a pragmatic and a theoretical point of view, the advantages of a statistical framework are persuasive:
• Sensor signals and associated features show a probabilistic behavior due to sensor noise, varying illumination conditions or segmentation errors.

• Pattern recognition routines should use all available sources of information including prior knowledge and empirical data.

• Probabilistic models give a unified mathematical formulation incorporating all modules.

• Decision theory guarantees the optimality of Bayesian classifiers, which maximize posterior probabilities.

• The design of learning algorithms can utilize comprehensive results in statistics and statistical learning theory.

• The success of probabilistic models in different areas of applied pattern recognition, such as speech recognition and handwritten character recognition, also motivate the use of statistical methods.

In addition to these general advantages, a probabilistic setting introduces some valuable tools for simplification and for the increase of computational tractability; the incorporation of independency assumptions regarding observed features leads to compromise solutions and paves the way to eliminate the trade-off, between computational efficiency and models that are still rich enough to provide the required discriminating power. Marginalization, that is the elimination of random variables by integration, reduce the complexity, allow the usage of probabilistic models, if the input data are incomplete, and provide techniques to define hierarchical modeling schemes.