Among other things, recognition and reconstruction or structure determination are the two important facets of Computer Vision. Recognition assumes knowledge about an object in the form of a model and using the measurements on the unknown object it determines the identity of the object from the knowledge base of models. On the other hand, reconstruction relies on the image features to determine the shape/structure of the object. A priori knowledge of the object in the form of a model is not essential in this case. However, images have to be taken under varying viewpoints to unravel the underlying structure of the object. If the knowledge about the surface is known i.e., whether it is a quadric surface or a planar surface etc. then it greatly helps in obtaining the structure parameters.

As recognition and reconstruction have wide applications in Robot Vision, Biomedical Imaging, Industrial Automation and scores of other applications, we are motivated to study the issue of recognition and reconstruction in the context of 3D objects. The recognition of 2D objects is quite simpler and many results are available in the literature. Earlier research was concerned with recognition of 2D or 2½D objects from 2D intensity images. Over a period of time, emphasis has been gradually shifted to recognizing 3D structures from 2D intensity images in one or more perspective views. Reconstructing an object from a sequence of images is called the structure from motion problem in the literature.

This thesis is devoted to the recognition of 3D objects from range images because they are insensitive to source of illumination, the nature of surface, occlusion, noise etc. and also concerned with the reconstruction of surfaces using some specific images and determination of depth from a sequence of monocular images in the case of nonavailability of range images.
1.1 LITERATURE SURVEY

1.1.1 Segmentation

Before trying to recognize the object, segmentation of range images is an essential step. Three types of approaches have been proposed in the literature: 1) Region based 2) Edge based 3) Combination of these two. A comprehensive review on image segmentation is given in [Pal and Pal, 1993]. Some contributions regarding this aspect are described in the following:

The covariance approach in [Berkmann and Caelli, 1994] segments the surface into planar, parabolic and curved regions as well as jump and crease discontinuities.

The method of Besl and Jain [1988] is based on Gaussian and mean curvatures whose signs yield eight surface primitives: peak, pit, ridge, saddle ridge, valley, saddle valley, flat (planar), and minimal. The surfaces are approximated into any one of eight surface types by using bivariate polynomials of order 4. The first stage of the algorithm creates a surface type label image based on the local information using mean and Gaussian curvatures. The second stage takes the original image and the surface type image as input and performs an iterative region growing using the variable order surface fitting.

Hoffman and Jain [1987] have used a clustering algorithm including methods based on minimal spanning tree, mutual nearest neighbour, hierarchical clustering, and the square error clustering of which the latter is found to be successful. Using the unit surface normal, a best fitting tangent plane is found. Then these patches are classified as planar, convex or concave. In the final stage, boundaries between adjacent surface patches are classified as crease or noncrease edges, and this information is then used to merge adjacent compatible patches to result in reasonable faces of the objects.

Yokoya and Levine [1989] have combined both region and edge based considerations. They form two edge maps: one for the jump edge and the other for the roof edge. These two
edge maps together with the curvature sign are then used to obtain the final segmentation.

Using a data structure of tessellation of image domain, combined with adaptive surface approximation technique, Schmitt and Chen [1991] accomplish a range image segmentation.

Chu and Aggarwal [1990] have integrated the segmentation maps obtained using region and edge segmentation maps as input. The result of integration is a region map in which each region is large and compact. Their algorithm accepts multiple inputs and applies user selected weights on various information sources. The source of integration is parametrically controlled for the desired spatial resolution. A maximum likelihood estimate provides initial solution of edge positions and strengths from multiple inputs. The edge map is converted into a region map using closed edge contours.

Biswas et al.[1995] have treated the segmentation problem as quantization problem. The surface normals are quantised to predetermined direction, i.e., surface regions are approximated to planar surface patches in some predetermined directions. The quantization is achieved through shift, subtract and threshold operations.

A generalized step edge detector is developed in [Ghosal and Mehrotra, 1992] to identify different kinds of edges in range images. These edge maps are thinned and linked to provide final segmentation. The parameters of the generalized edge model, viz., orientation, two slopes, one step jump at the location of the edge and background grey level, are determined using Zernike moment based masks.

Data aggregation is performed via model recovery in terms of variable order bivariate polynomials using iterative regression [Leonardis et al., 1995]. Model recovery is initiated independently in regularly placed seed regions in the image. Selection of models is defined as the quadratic Boolean problem and solution is sought by the WTA (Winner-Takes-All) technique.
Cung et al.[1990] have proposed a multiscale representation in terms of well localized depth and orientation edges for segmentation. The extraction is accomplished first by detecting the presence of significant edges at coarse scale and then determining their location by tracking them over decreasing scale. A multiscale thresholding is used to inhibit the insignificant details. Then edges are classified into "true edge" and "diffuse edge" by invoking classification rules derived from mathematical analysis of edge displacement and branching over scale space.

A robust method for registration and segmentation of multiple range images is proposed in [Masuda and Yokoya, 1995]. The registration algorithm determines rigid motion parameters from a pair of range images. The segmentation method classifies the input data into four categories comprising inliers and 3 types of outliers. The inliers obtained from multiple range images are integrated to construct a data set representing an entire object. In this thesis an edge based approach is followed.

Having discussed some approaches to segmentation, the next phase is concerned with the representation of segmented parts for the purpose of recognition.

1.1.2 Object Recognition

The Object Recognition systems are designed to perform a variety of tasks in the factory environment such as inspection of parts and sorting of products into bins. The recognition system compares a scene with entities in a model base containing a description of each object the system is expected to recognize. Designers of such systems must choose a representation strategy for their object models. Additionally, the nature of sensed image viz., intensity or range, affects the recognition strategy. The choice of representation strategy and the image type results in the following four types of recognition systems.
(i) **2D - 2D (2D images, 2D model)** - Matching strategies are often practical and implemented under certain constraints.

(ii) **2D - 3D (2D image, 3D model)** - Here, the descriptive power of 3D models may be combined with sensing strategies.

(iii) **3D - 2D (3D image, 2D models)** - This is wasteful in the sense of ignoring depth information from the image.

(iv) **3D - 3D (3D image, 3D model)** - These have a wide scope and are discussed extensively in the literature.

Some of the representation schemes which are in use today are:

(a) **Wire frames**: A wire frame representation is a graph whose vertices are 3D points on the object surface and graph edges represent a physical edge on the object. These representations do not contain the surface information.

(b) **Constructive solid geometry (CSG)**: In this modelling, a finite number of primitive shapes (spheres, cylinders, etc.) are combined using boolean operations to produce the desired result. A CSG model is stored as tree with leaf nodes representing the primitive solids, internal nodes representing the regularized boolean operations and arcs enforcing precedence among operations.

(c) **Surface models**: The surfaces can be represented in a variety of ways. Three of the most popular approaches are:

- the implicit surface form

\[ S = \{(x,y,z): f(x,y,z) = 0\} \]

Ex: Planar surface, quadric surface

- the general parametric form

\[ S = \{(x,y,z) : x = f_1(u,v), y = f_2(u,v), z = f_3(u,v)\} \]
(u,v) \in D \subseteq \mathbb{R}^2} \}, \text{ and} \\
\text{Ex: B-spline surfaces} \\
\bullet \text{the graph surface form} \\
S = \{(x,y,z) : z = f(x,y), (x,y) \in D \subseteq \mathbb{R}^2\}

Clearly, the graph surface is just a special case of the general parametric surface. For example, range data is usually viewed as samples from a graph surface.

(d) **Volumetric models**: Here the object is subdivided into small "primitive" volume elements called voxels with reference to some coordinate system. Some of the volumetric descriptions are cubes, superquadrics and hyperquadrics.

(e) **Sweep representations**: The object volumes are swept by the generating curves with respect to one or more axes. In vision community, the well known sweep representation is Generalized cylinder representation.

(f) **Geons**: These are qualitative representations of any object. Each geon has four qualitative features:

\begin{itemize}
  \item Edge: straight or curved;
  \item Symmetry: rotational/reflective, reflective, assymetric;
  \item Size variation: constant, expanding, expanding/contracting;
  \item Axis: straight or curved.
\end{itemize}

(g) **Multiview representations**: In this representation a 3D object is described by some or all of its possible 2D projections which are often known as 3D or view centred models. Other representation methods attempt to combine all the viewpoint specific models into a single data structure.

Some of popular paradigms for object representation are interpretation tree search, evidence based recognition, graph matching, geometric hashing, automatic
A number of systems for 2D or 3D object recognition have been developed. Typical examples are 3DPO by Bolles and Horaud [1986], ACRONYM by Brooks [1983], BONSAI by Flynn and Jain [1991], INGEN by Vayda and Kak [1991] etc.

A brief survey of various techniques employing different object representations and matching strategies is given in the following:

The approach of Zhou and Davis [1990] aims at increasing the efficiency of tree search using heuristic and relaxation techniques. A structural description of 3D object recognition is proposed.

A 3D object is represented by the linear combinations of 2D images of the object in [Ullman and Basri, 1991]. The approach handles the rigid 3D transformations of the object, sharp and smooth boundaries of the object.

Using the geometrical facts about bitangencies, creases and inflections, the descriptions of the surface are obtained from the image outlines. These descriptions are unaffected by the viewpoint or the camera parameters. The surfaces that are dealt with have a rotationally symmetry [Forsyth et al., 1992].

A View Independent Relational Model (VIRM) is designed by Zhang et al.[1993] to recognise known 3D objects from monochromatic images of unknown scenes. The system inspects the CAD model from different viewpoints and a statistical inference is applied to identify relatively view independent relationships among the component parts of the object. These relations form the nodes in the hypergraph. During recognition, visible features are matched to the nodes of the hypergraph.

By incorporating view-independent features and the object centred representation, a compact object representation is provided in Hwang at al.[1993]. The solid modeller is based on CAD. The matching process is based on the constrained search paradigm.

The recognition of partially occluded 3D object is considered by Ray and Dutta
Majundar [1994]. The recognition uses principal curvature, mean Gaussian curvature as the local descriptors (invariant features) of the surfaces. Object models are generated from canonical perspectives and the feature matching is realized through Hopfield neural networks.

Wang and Cohen [1994] have identified the shape of 3D object from image curves using B-spline curve modelling. First, 3D control points are found from the corresponding curves in each image in stereo imaging system. A Bayesian framework is used for classifying the image into one of possible surfaces based on the extracted 3D object curves. The object curves are unwarped into planar curves before the matching process.

A neural network approach to Constructive Solid Geometry (CSG) based 3D object recognition is proposed in Chen and Lin [1994]. Precedence graphs in which each node represents a primitive volume and each arc between the nodes represents the relation between them are used for object representation. Object recognition is achieved by matching the scene precedence graph to the model precedence graph. The matching constraints include match validity, primitive similarity, precedence graph preservation and geometric structure preservation.

An Invariant Feature Indexing (IFI) based object recognition system is designed, which utilises multiple feature group and multiple-view information to reduce the number of hypotheses at the verification stage [Mao et al., 1995]. Out of the two approaches proposed, the first one tallies the number of consistent votes cast by the prototype hypothesis for particular object models and the second one examines the consistency of estimated object pose from multiple scene-tuples in one or more views.

Probability indexing is suggested by Olson [1995] in order to restrict the number of possible matches between sets of image features and sets of model features. The indexing uses sets of three points undergoing 3D transformations of projection by taking advantage of the probabilistic peaking effects.
As the individual cues are fallible and often ambiguous, a multiresolution system that integrates perceptual organization (grouping), segmentation, stereo, shape from shading and line labelling modules is proposed in [Pankanti and Jain, 1995]. The output of the integrated system is shown to be insensitive to the constraints imposed by the individual modules and produces better reconstruction of the input scene than the individual modules.

Shape of an object is described by a prototype template which consists of the representative contour/edges and a set of probabilistic deformation transformation on the template. A Bayesian scheme is employed to find a match between the deformed template and the objects in the image. The proposed scheme [Jain et al., 1995] is invariant to location, rotation and moderate scale changes of the template.

The recovery of 3D shape information from 2D images by two networks is given in [Ardizzone et al., 1994]. The first extracts a brightness gradient map from the image and the second which is a propagation network estimates the geometric parameters of the object parts present in the acquired scene.

The surface is parameterized by defining a one-to-one mapping from the surface of the original object to the surface of the unit sphere [Brechbuler et al.]. The initial mapping is based on conduction model. The parameterization enables to expand the object surface into a series of spherical harmonic functions which extend the 3D concept of elliptic Fourier descriptors to 2D closed curves.

In this thesis, the quadric surface representation is used for both scene and the models.

1.1.3 Object Reconstruction

Constructing an object from its sensed images has potential applications in Biomedical engineering, Robotics and Multimedia. Work on surface reconstruction began with one or two views in the literature and was later extended to include several views as the
simple solutions to correspondence problem emerged. The later category comes under the realm of structure from motion. The correspondence is confined to tracking features, viz., points, lines and curves and is then used to estimate either the structure or motion parameters or both. For example, solution to eight point correspondence problem leads to the essential matrix which is decomposed to yield motion parameters. Knowledge of motion parameters helps to find the structure parameters. Presently we do have approaches that estimate optimally both the structure and motion parameters in one go.

Some of the approaches dealing with reconstruction of object from one to three views are described in chapter 4 and those dealing with structure from motion, i.e., from a sequence of images, are relegated to chapter 5.

1.2 THE ORGANIZATION OF THE THESIS

The segmentation of range images, representation surfaces by quadrics and recursive matching technique employing the Kalman filter are presented in Chapter 2. Shortlisting of a model that fits the scene in the least squares sense from a set of models represented in terms of contour data is proposed in Chapter 3 using Kalman filter based contour matching technique. Chapter 4 gives the methodology for the determination of quadric type structure for the surface using 3 static views. It also presents the computation of qualitative information about the surface. Chapter 5 gives a method for the determination of depth from a sequence of views using Kalman filter based formulation. Finally, the summary of the results and suggestions for further work are relegated to Chapter 6.