CHAPTER 1

PRELIMINARIES

1.1 OUTLINE OF THE THESIS

The thesis aims to discuss graph cuts and its various models applied to variety of image processing problems. The *first chapter* takes care of the preliminaries related to the work presented in the thesis. There are two areas the work presented in the thesis is closely related with. Graph theory and Computer vision or in general image processing. Essential preliminaries related to both the areas are briefly mentioned in the first chapter.

The *second chapter* briefly explains how most of the image processing problems under consideration can primarily be considered as image labeling problems. It mainly focuses on the optimization approach to handle labeling problems. Standard form of the objective function used to address labeling problems via optimization approach involves two terms: Data term and structural term. Uniformly smooth structure, segment- wise constant structure and segment-wise smooth structure and related forms of objective functions are discussed. A brief review of two main types of optimization techniques (Global and local) is presented.

In the *third chapter*, popular graph cut models are discussed. There are mainly three types of models depending on structural term of the objective function. At the beginning of the chapter, a graph cut model for universally smooth structure is discussed. Two other graph cut models for segment wise smooth structure are elaborated. The first model optimizes the objective function using interchange moves whereas the second addresses the optimization problem using growth moves. Graph cut model using shift moves to handle universally constant structure is presented at last in the chapter.

The main objective of the research work is to explore the concept of Graph cuts thoroughly and to apply it to newer problems. We tried to study a problem of binarization of textual images and developed a simple graph cut model to address it. Implementation of the model is carried out with Java programming language. The *forth chapter* presents the mathematical details of the model along with results of computation.

Graph cuts basically assign two values to objects under consideration efficiently (through energy minimization concept) in single iteration. There are varieties of problems, not limited to computer vision, which are or can be addressed by optimization. We tried to explore, which types of such problems could be addressed by graph cuts. In *chapter five*, we studied what kind of objective functions can be dealt with by graph cuts. Characterizations of two classes (O^2 and O^3) of objective functions minimizable by network flow terminology are developed and studied.

1.2 PRILIMINARIES

In this chapter, the prerequisites for the research-work are briefly discussed. Graph theory and computer vision or broadly image processing lays the foundation of the work presented in the thesis. The first section discusses the important components of Graph theory, to be more specific, network flow terminology and associated algorithms, which are essential for the work of the thesis. The second section is devoted to components of computer vision essential for the work of the thesis.

1.2.1 GRAPH THEORY

To start the introductory discussion of graph theory, we first begin with some terminology. First of all, we should start with the definition of graph itself. A graph *G = (V, E)* is made up of two sets, a set of vertices $V = \{v_1, v_2, ..., v_n\}$ and set of edges $E = \{e_1, e_2, ..., e_m\}$. An edge is a connection between one or two vertices. An edge e_i is an unordered pair (v_j, v_k) of vertices $v_j, v_k \in V$. If $v_j = v_k$, then e_i is called a loop. To illustrate the concept, we consider the graph shown in Figure 1.1.

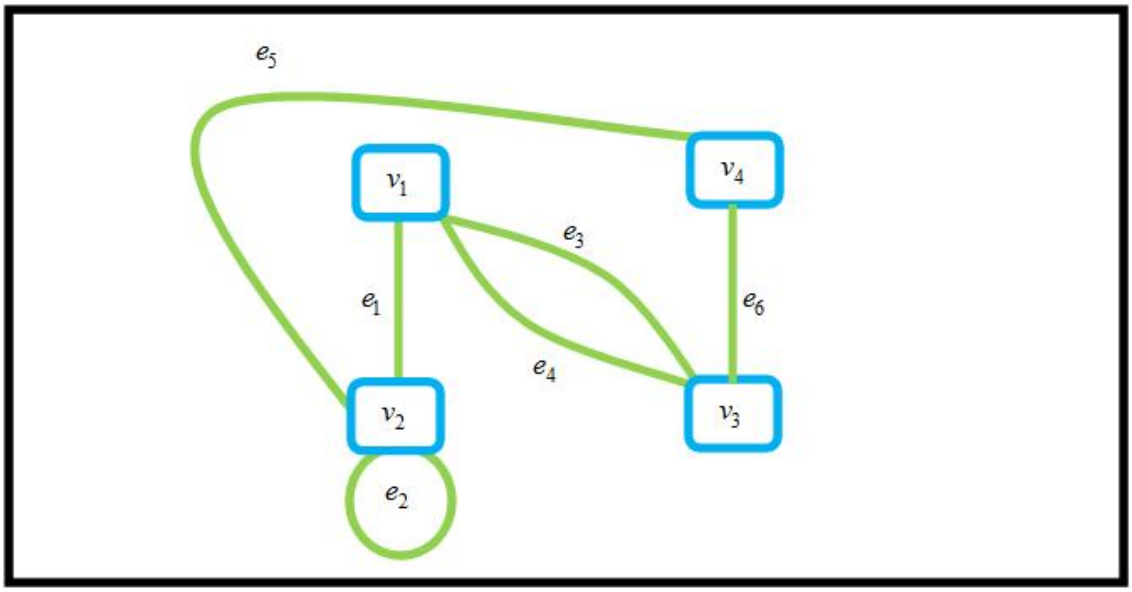

Figure 1.1: Graph with parallel edges and loop

If we denote this graph by G, then *G = (V, E),* where $V = \{v_1, v_2, v_3, v_4\}$ and $E = \{e_1, e_2, e_3, e_4, e_5, e_6\}$. Note that, $e_1 = (v_1, v_2)$, $e_2 = (v_2, v_2)$, $e_3 = (v_1, v_3)$, $e_4 = (v_1, v_3)$, $e_5 = (v_2, v_4)$ and $e_6 = (v_3, v_4)$. Here, edge e_2 is a loop as it connects v_2 with itself.

This type of graph is called *undirected graph*, since edges do not have direction. *Directed graph* is a graph in which edges are ordered pair of vertices. If the edges are defined to be ordered pairs of vertices, the graph is called a *digraph*. In directed graph, an edge $e_i = (v_j, v_k)$ represents edge from vertex v_j to v_k. In case of digraphs, $(v_j, v_k) \neq (v_k, v_j)$. Here, we have considered a graph, where edges do not have weights. If edges are assigned corresponding edge weights, then the graph is called *weighted graph*. If the weight of an edge $e = (v_j, v_k)$ is *c*, then we write, *w* (*e*) = $w(v_j, v_k) = c$ or sometimes $|e| = c$. A path between two vertices v_1 and v_n is an ordered sequence of vertices $v_1, v_2, ..., v_n$, where every pair of consecutive vertices are connected by means of an edge. A path from v_1 to v_n can also be defined as a sequence of edges $e_1, e_2, ..., e_{n-1}$, where the first edge e_1 is incident on v_1 and v_2, the second edge e_2 is incident on v_2 and v_3, and so on, the last edge e_{n-1} is incident on v_{n-1} and v_n. Note that, in a case of directed graph or digraph, a path $v_1, v_2, ..., v_n$ from vertex v_1 to v_n need not necessarily be a path from v_n to v_1 as well, since $(v_i, v_{i+1}) \neq (v_{i+1}, v_i)$. Two

vertices v_1 and v_2 are said to be *path-connected*, if there is a path from v_1 to v_2. A graph is said to be *connected*, if every pair of its vertices are path-connected. The graph, which is not connected, is called *disconnected*.

A *cut* is a minimal collection of edges removal of which from the graph makes the graph disconnected. For the graph given in figure 1.1, there are various possible cuts. Some of the cuts are $\{e_1, e_6\}$, $\{e_1, e_5\}$, $\{e_3, e_4, e_5\}$ and $\{e_5, e_6\}$.

A *flow network* is a directed graph *G = (V, E)* in which every edge $(u,v) \in E$ has a *capacity* $c(u,v) \geq 0$. It is customary to refer origin and destination vertices in a flow network as the *source* and the *sink,* respectively. We assume that, in a flow network *G = (V, E)* with source *s* and sink *t*, there is a path from *s* to *t* that passes through *v,* for any vertex $v \in V$.

A *flow* in a flow network *G = (V, E)* with source *s* and sink *t* is a function $f : V \times V \rightarrow R$ that satisfies the following properties:

1. *Capacity Constraint*: For all $u, v \in V, f(u,v) \leq c(u,v)$.
2. *Skew Symmetry*: For all $u, v \in V, f(u,v) = -f(v,u)$.
3. *Flow Conservation*: If $u \in V$ and $u \neq s, u \neq t$, then, $\sum_{v \in V} f(u,v) = 0.$

We say that, $f(u,v)$ is a flow from vertex *u* to vertex *v*. For two disjoint sets *X* and *Y* of vertices and a flow function *f,* we define $f(X,Y) = \sum_{x \in X} \sum_{y \in Y} f(x,y)$ and $c(X,Y) = \sum_{x \in X} \sum_{y \in Y} c(x,y).$

The value of a flow *f* is denoted by $|f|$ and defined as $|f| = \sum_{v \in V} f(s,v)$; i.e. the total flow out of the source, which is same as $\sum_{v \in V} f(v,t)$, the total flow gathering at sink *t*. The network shown in figure 1.2 is an example of a valid flow network.

In the Figure 1.2, the flow network *G = (V, E)* with vertices $V = \{s, v_1, v_2, v_3, v_4, t\}$ and edges of edge set *E*, where the capacity $c(u,v)$ of each edge (u,v) is mentioned on the edge. In figure 1.3, the flow across each edge of a flow function f is labelled to the left of each edge's capacity. For example, the entry 1/5 mentioned for edge (v_4, v_2) indicates that, the edge has a flow of 1 unit and capacity 5 units. It should be noted that, $f(v,u) = -f(u,v)$ for all $(u,v) \in E$. It can be easily verified that, *f* satisfies the capacity and the skew symmetry constraint. By summing the flows into and out of each vertex, it follows that, *f* also satisfies the flow conservation properties of flow function. The value of the flow *f* in this case, is $|f| = 7$, which is the total flow out of the source *s* and it is same as the total flow into the sink *t*.

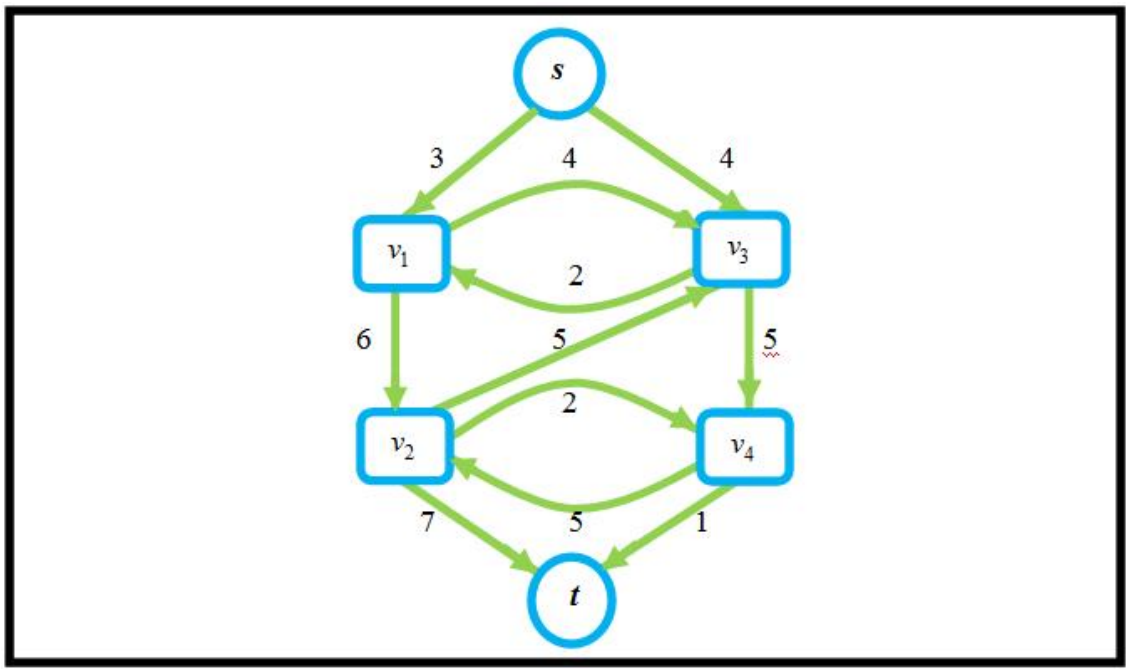

Figure 1.2: A flow network G = (V, E) where each edge $(u,v) \in E$ *is labelled with* $c(u,v)$

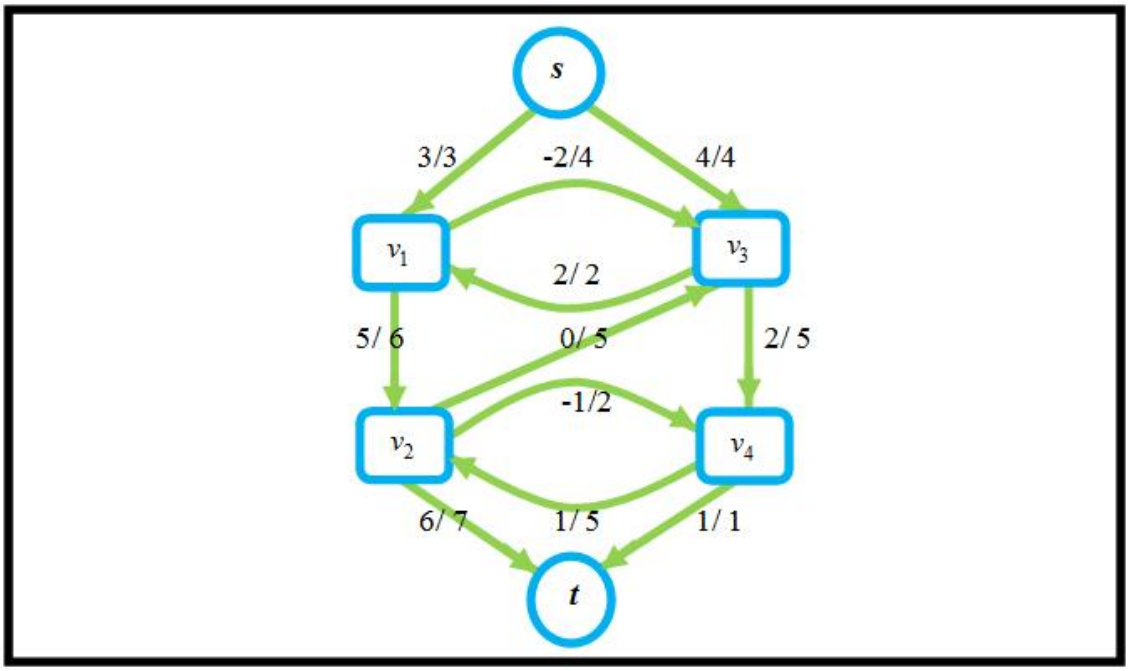

Figure 1.3: the same flow network G = (V, E) where each edge $(u,v) \in E$ *is labelled with* $f(u,v)/c(u,v)$ *for a flow function f*

Additionally, for a flow network *G = (V, E)* and a flow *f*, we define *residual capacity* of an edge $(u,v) \in E$ to be $c_f(u,v) = c(u,v) - f(u,v)$; i.e., the difference between the capacity of the edge and the flow currently being sent across an edge. The *residual network* of *G* induced by the flow *f* is $G_f = (V, E_f)$, where $E_f = \{(u,v) \mid (u,v) \in E \text{ and } c_f(u,v) > 0\}$. The residual network G_f of the

flow network given in above example is pictured in Figure 1.4 where, each edge (u,v) is labelled with $c_f(u,v)$.

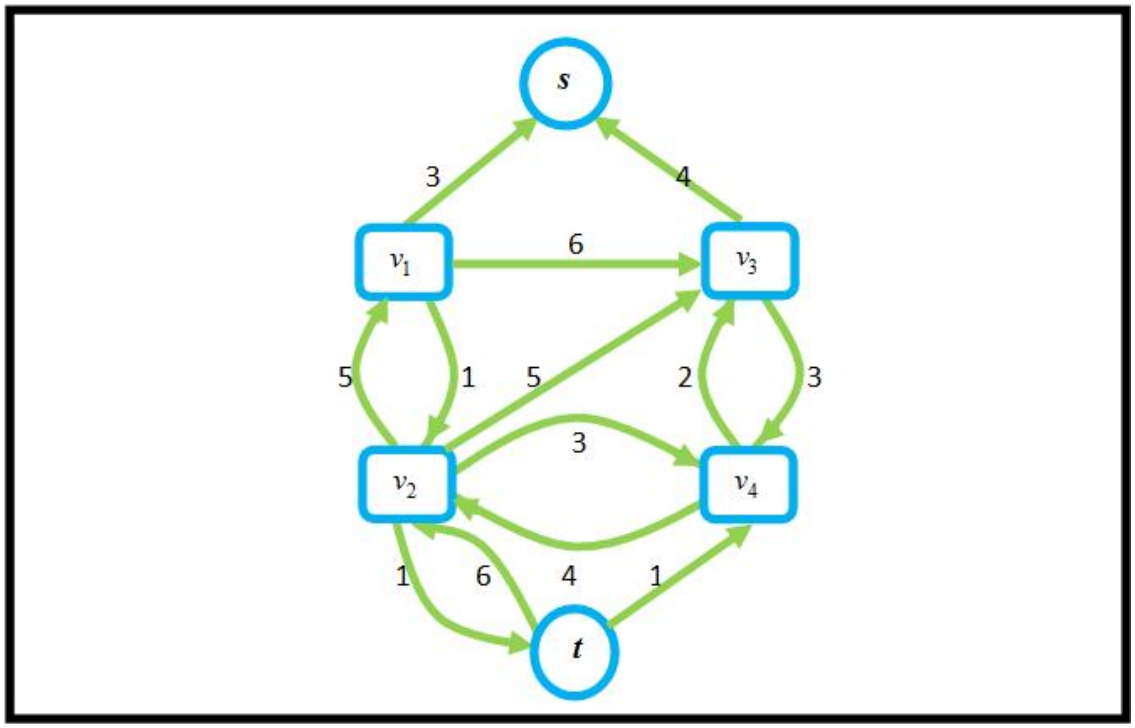

Figure 1. 4: The residual network G_f of the flow network pictured in Figure 1.3

For a flow network *G = (V, E)* with source *s* and sink *t*, the question 'what flow function *f* maximizes $|f|$?' turns out to be important. Ford and Fulkerson proposed an algorithm to address the problem. The algorithm is popularly known as Ford- Fulkerson algorithm.

1.2.1.1 FORD FULKERSON ALGORITHM

Let *G = (V, E)* be a flow network with source *s* and sink *t*. The aim is to define a flow function f that maximizes $|f|$. Note that $c_f(P)$ is simply a temporary variable for the residual capacity of the path P.

1. For each edge $(u,v) \in E$, $f(u,v) = f(v,u) = 0.$
2. If there is a path P from source *s* and sink *t* in the residual network G_f continue the step3. Otherwise terminate.
3. Set $c_f(P) = \min_{(u,v)\in P} c_f(u,v)$.
4. For each $(u,v) \in P$, set $f(u,v) = f(u,v) + c_f(P)$ and $f(v,u) = f(v,u) - c_f(P)$.
5. Return to step 2.

The Ford – Fulkerson algorithm starts with zero flow along the given flow network. It searches for augmenting paths. If an augmenting path P is found, it works out $c_f(P)$, the maximum value of flow that could be pushed through the path P and update the flow along P accordingly. The algorithm keeps on searching for augmenting paths and repeat the procedure until there is no augmenting path in the flow network. The resulting flow network has the maximum possible flow.

1.2.1.2 MAX- FLOW MIN-CUT THEOREM

Let $G = (V, E)$ be a flow network with source s and sink t, and flow function f, and let G_f be the residual network of G induced by f. Then the following statements are equivalent:

1. f is maximum flow in G.
2. G_f has no path from s to t.
3. $|f| = c(S,T)$ for some cut (S,T) of G.

1.2.2 COMPUTER VISION

Digital image can be viewed as a rectangular arrangement of pixels, where pixels are a fundamental blocks which carry intensity. In the RGB image, every pixel represents a triplet of intensities corresponding to the three fundamental colours red, green and blue. In grey scale image, every pixel represents its grey value, which happens to lie in the range of 0 to 255. In mathematical terms, we can consider an $m \times n$ greyscale image as a matrix carrying mn entries, where every entry is a grey value, which can be any no. from the set $\{0,1, .., 255\}$. A pixel with intensity value zero is completely black in colour, whereas the one with intensity 255 is of complete white colour. Throughout our study, we are going to work with grey scale images. However, the results and the mathematical models presented in the thesis are not restricted to grey scale images. An image other than grey scale can be first converted to grey scale and then the models can be applied to it.

Computer vision is a branch of information technology which deals with acquisition, analysis and interpretation of images and sometimes even, drawing inference from the images. In simple terms, computer vision strives to enable machines to draw conclusions from the images. Actually, it's more about imitation of human skills of vision. Humans can easily interpret the images and can make fruitful inferences based on it. The aim of computer vision is to empower the computers with similar capabilities. However, the task is not as easy for computers as it is for human beings. Even the simple task of identifying and differentiating between the objects lying in the image, which is a very basic but fundamental task of human vision system, turns out to be non-trivial and considerably complex for the machine. Identification of the characters of the scanned script or of a handwritten document by the machine turns out to be even more difficult computer vision problem due to variation in either the handwriting of different individuals (in case of handwritten document) or fonts and font-sizes of the document (in case of printed textual documents). Thus, understanding the images by extracting the hidden information within it using variety of models involving statistics, mathematics, physics and learning theory is the objective of computer vision and broadly of image processing.

1.2.2.1 RECOGNITION

Computer vision deals with a vast range of problems ranging from image segmentation to facial recognition and pose estimation. We present a brief overview of the fundamental computer vision problem called object recognition or simply recognition.

As mentioned in the introduction, recognition of the objects of the image is the crucial aspect of image interpretation and understanding. In a simple picture, there can be variety of objects. Human vision system can easily identify and differentiate between the objects present in the image. Note that, the object (for example, family members in a family photograph) present in the image can vary in shape, size and angle of projection etc, than too, human vision system can easily identify it. The same

task achieved through machines can lead to some standard recognition problems. Few of the recognition problems are object recognition or classification, object identification and object detection.

On the basis of the pre–specified object, identification of all instances of the objects present in the image can be made. i.e. all segments of the image representing the pre-specified object can be identified and their location in the image can be deduced. Similarly, in a video sequence, a pre-specified object can be identified. There are numerous approaches to handle the problems of object recognition. There are mainly two types of approaches: 1) Appearance based approach and 2) Feature based approach.

In appearance based methods, exemplars of the object to be recognized are used for the task. Note that, objects to recognize may look different under different lighting conditions, from different angles and in different sizes. Thus, a single exemplar is not sufficient to guarantee accurate results. Divide and conquer search, edge matching, matching using greyscale, gradient matching, Histograms of receptive filed responses, large model bases are few of the techniques relying on appearance based approach.

1.2.2.2 APPEARANCE BASED METHODS

In *edge matching approach*, the edges of objects and image are detected and compared. As the edges are very unlikely to change due to change in colour, shape or size of the images, this approach handles object recognition problem very well. Counting the no. of overlapping edges turns out to be a good strategy, but it can't handle the change in shape efficiently. However, probability distribution of the distance of nearest edge in search image provides the best results.

In *divide and conquer search approach*, the set of all positions of the cell is used. The least possible value of best position is evaluated. If the value is too large, the cell is trimmed. If the value isn't happened to be too large, the cell is divided into smaller sub-cells and the method give the same treatment to all the sub-cells recursively until the process culminates in the smaller enough cell. In case of precise evaluation of least values, the process guarantees the determination of all the matches fulfilling the criterion. However, sometimes evaluation of the least values can be troublesome.

Although edges are invariant under change of brightness, considering only edges for the recognition leads to loss of lots of information, which prevents edge matching approach being the best approach. The approach which is invariant under the change in lighting or colour and also preserves the information which is missed in the edge detection is *gradient matching*. The approach is similar to grey scale matching. In *grey scale matching*, pixel distance is represented as a function of pixel position and pixel intensity. This approach is also applicable in case of coloured images.

The *approach based on histogram* (histogram of receptive field responses) discourages the unnecessary correspondence of points. Receptive field responses encode correlation between different image points. In the *large model bases*, the eigenvectors of the exemplars i.e. eigenfaces are compared with the image segments. In fact, eigenfaces are the geometric models of the segments to be recognized.

1.2.2.3 FEATURE BASED METHODS

In feature based methods features of image and object to be recognized are matched. However, single position in the object must justify all possible matches. In this approach, features like surface patches, corners and direct edges of both the object and the image are identified. Feature based methods include interpretation trees, hypothesize and test, pose consistency, pose clustering, invariance, geometric hashing, Scale-Invariant Feature Transform (SIFT) and Speeded Up Robust Feature (SURF).

The method of *Interpretation trees* is a historic one, which is currently not used widely. In this method the search is made through trees. Each vertex in the tree leads to a set of matches. The root vertex does not lead to any set, or in other words, it leads to an empty set, whereas all other nodes give union of matches of parent vertex and additional match. When the set of matches is infeasible, nodes are removed. The deleted vertex has no children.

In *correspondence and hypothesize method*, the correspondence between object features and image features are hypothesized, which is used to generate fruitful hypothesis about projection from the object coordinate frame to image frame. This hypothesis is then used for back-projection. If the result of back-projection and image are matching almost everywhere, the hypothesis is accepted. There are numerous ways to generate hypothesis. The hypothesis for object is equivalent to hypothetical location and direction when camera parameters are known. Small sets of object features are correlated to appropriate segments of the image. There are three main approaches for hypothesis: 1) by pose consistency 2) by pose clustering 3) by invariants. Randomization and grouping are also important for the search. For randomization, set of image feature are scrutinized, until the probability of the missing object turns negligible.

Note that, $(1-F^o)^n = P$

Where,

F = the ratio of 'good' image points,

o = the number of correspondences necessary,

n = number of trials,

P = the chance of individual trial given that at least one correspondence is incorrect

In the method *of pose consistency*, the object is aligned to the image, that's why the method is sometimes also called *alignment*. There is a geometric constraint – if the image features and object features matches almost everywhere, it is very unlikely to be due to merely coincidence. Recovery of unknown camera parameters can easily be made in case of matching of large no. of image features and object features.

Pose clustering essentially uses a Hough transform. All segments of image corresponding to the object have negligible difference in their poses. Accumulator array of every object describes its pose space. Hypothesis of correspondence between image frame group and every frame group of every object is formulated. High correspondence in the accumulation array of the object indicates the presence of the object at that pose.

Invariance approach depends on the geometric properties which are invariant to camera transformations. The technique is mainly used for planer objects but can also be applied to non-planner objects, subject to few modifications. *Geometric hashing* is a technique similar to pose clustering, but rather than focusing of correspondence of pose as in pose clustering, geometric correspondence is addressed in this technique. The method is frequently used for Computer aided design and medical imaging.

Speeded up robust features (SURF) is a strong image detector and descriptor which relies on sums of approximated two dimensional Haar wavelet responses. The technique is proved to be time efficient when compared to other methods.

Apart from the appearance based approach and feature based approach, there are other approaches which can also deal image segmentations. For example, genetic algorithm, Biologically inspired image segmentation, Artificial neural network and deep learning, Explicit and implicit three dimensional objects, Fast indexing, global scene representations, gradient histograms, unsupervised learning, Bingham distribution, etc.

Till now, we have talked about a specific problem of object recognition or object classification. Sometimes recognition problems appear as identification or detection problems. Object identification problems include identification of a particular individual's face, fingerprint, etc. Identification of characters in handwritten scripts, identification of a specific vehicle in a surveillance system, etc.

1.2.2.4 PRACTICAL PROBLEMS BASED ON RECOGNITION

There is a variety of special problems based on recognition. *Image retrieval based on content* is a special application of recognition problem. Given a large collection of images and content, machine should identify all the images from the given collection which belong to the target group. e.g. , from the collection of images, identification all the images consisting cars, or more specifically, consisting all images containing cars of a particular model is an example of content based image retrieval.

Pose estimation is also one of the most important applications of recognition. The orientation of the object with respect to camera can be estimated by treating it as a recognition problem, which is practically important in case of automated production line system, where automated system has to identify the object moving on the assembly line and has to remove it from the conveyer belt. For the task, exact location and orientation of the object are the crucial information for the machine as the machine has to catch the object from the conveyer belt before moving it from the belt. The task is complex as the object keeps on changing its position due to moving conveyer belt.

Optical character recognition is a problem which can be treated as recognition problem. The problem is to identify the characters of the scanned textual document where text is either machine typed or is a handwritten script, so that, the text can be converted into the format, where editing of the text becomes possible. This problem is of special importance for us as we have worked on the sub-problem of this task, which is binarization of the text image. We will discuss this special problem of recognition once again, with detail in the later chapter.

Facial recognition is another important practical problem, which is handled by recognition. In this problem, from a single image or from the set of images or from a video clip, to identify face of individual or individuals is the motive of the problem. This problem is currently handled by almost every digital camera available in the market, and thus, is a popular problem based on recognition.

Shape Recognition technology is a technique in which humans are identified from the image consisting of various objects along with humans. The technique is crucial for automated surveillance systems. The system uses head and shoulder patterns for discriminating humans and objects.

1.2.2.5 SEGMENTATION AND OTHER COMPUTER VISION PROBLEMS USING APPROACH OF OPTIMIZATION

One can easily find that, the goal of computer vision is too ambitious, as the information present in the two dimensional images gives only a glimpse about the real situation in the three dimensional world. Some of the very important problems of computer vision like image segmentation, image restoration, visual correspondence etc. can primarily be defined as image labelling problems. But, the complexity of the problem increases due to uncertainty associated with the imaging process, and it leads to many solutions. The issue is to find out the most appropriate solution associated with the problem. The *optimization approach* is generally being used to deal with the problem. An objective function is designed to measure the quality of the solution with reference to the type of the problem and constraints associated with it. Objective function of the problem is a real valued function that is defined on the set of all possible solutions. Note that, objective function may not be unique. But, whichever objective function one chooses, it must take care of constraints of the problem under consideration. There are two main constraints that the solution of the labelling problem has to satisfy. (i) The solution must be compatible with the available data. (ii) The solution must be in accordance with the prior or structural knowledge. The two constraints will now onwards be referred to as Data constraint and Structural constraint. On the basis of how well does the particular solution give justice to both the constraints, objective function measures the quality of the particular solution. In simple terms, the objective function measures the inappropriateness of the solution to the problem under consideration. i.e. lesser is the value assigned by the objective function to the solution, more appropriate it is as a solution of the problem. Thus, mathematically the labelling problem to be dealt can be considered as an optimization of objective function. However, the objective function can have many local minima; in other words, the objective function we want to minimize may not necessarily be a convex function, which increases the complexity of the problem.

The task of minimization of the objective function in most of the cases turns out to be computationally expensive and that's why people targets to get the approximate solution. Sometimes, even selection of objective function from the available choices is made on the basis of the computational complexity of its minimization. Thus, the formulation of good mathematical model to handle the labelling problem requires an art to balance the two tasks: (1) selection of appropriate objective function translating both the constraints of the problem (2) Choosing the proper minimization technique which leads to the (exact or approximate) global minimum of the objective function at reasonably low computational cost. However, the failure of the model puzzles the reasoning as the failure could have arisen due to poor selection of objective function or due to wrong choice of minimization technique.

One may naturally wonder why to go for minimization approach for vision problems if it ends up with this much complexity. The answer lies in the pros of the approach: (1) it provides a common structure for all similar vision problems. (2) The approach can be theoretically rationalized by Baysian statistics. (3) The approach allows us to take care of all constraints of the vision problem and the solution of the problem obtained by this method turns out to be the most appropriate solution with reference to the constraints. (4) This approach standardizes the process of solving different vision problems converging to labelling problems. i.e., after the selection of appropriate objective function,

different vision problems can be addressed using the same mathematical model and the same computational framework.

The biggest disadvantage of the approach is the computational complexity of it. In most of the cases, finding the exact minimum of the objective function is NP hard even in case of simple vision problems. However, the approximate solutions provided by the model are quite closer to the exact global minimum of the objective function and hence allows us to apply the approach in vision problems and exploit the plus points of the approach.