# 1
# Introduction

In today's time, despite the increasing trend in the creation and processing of digital data, handwritten or paper based storage is still active and growing [Hilbert and López, 2011]. This has in turn led to a growth in the research and applications in Document Image Analysis (DIA) systems. The primary purpose of any DIA system is, automated storage, extraction, and processing of the information present in documents. The research in DIA lies at a very interesting intersection of individual fields like pattern recognition, image processing, machine learning, artificial intelligence, storage systems and linguistics and encompasses several sub research areas, like document categorization and understanding, hand-written and typed text recognition, graphics recognition and graphics retrieval. Figure 1.1 depicts a typical example and various types of tasks [Kasturi *et al.*, 2002] that fall under the umbrella of DIA.

Graphics recognition is a sub-domain in Document Image Analysis which aims at analyzing the content of graphical documents. A graphical document can be defined as a virtual or a physical object created to convey a piece of information. This piece can manifest itself in a picture or a diagrammatic form, conforming to a graphical language. Examples of graphical documents can be engineering drawings, flowcharts, maps, diagrams, architectural floor plans, etc. Graphical languages are human-made communication systems enabling the expression of complex semantic concepts via graphical information. In graphical languages, the vocabulary comprises of a symbol set, and the grammatical rules are context based relations between the symbols in the symbol set which provides the complete meaning to the given graphics in the document. Thus, the understanding of a graphical document comprises of both parsing of, the unique, isolated symbols as well as the relationship between their context. Still, graphical languages, have variations in terms of, the nature of reader, the difficulty level in the comprehension of the enclosed information, design and creativity matters, etc. Apart from recognition of graphical symbols in a graphical document, the next important thing is retrieval of such documents based on the semantic content derived from the symbols identified in those documents.

The amount of data available in digital format has grown exponentially in recent years. Data, specifically in the form of images are generated in large quantities and stored every day, and this results in a massive, distributed, and in most cases unstructured multimedia information. There are mainly two ways using which one can access images from a database; they are:

1. Text based retrieval: In this case, all the images are tagged using manual or semi-automatic tools during the cataloging stage. Indexes are formed over the database based on these tags. After that a suitable data structure is created that suits the retrieval task. During the retrieval stage, traditional text based retrieval techniques are applied to get the result based on the user query.

2. Content-based retrieval: In this type of retrieval method the search is performed by specifying the actual content of the image data in terms of query image, sketch etc. For example, a Content Based Retrieval System, which is described as the process of retrieving desired images from a large collection on the basis of syntactical image features. Such a system uses

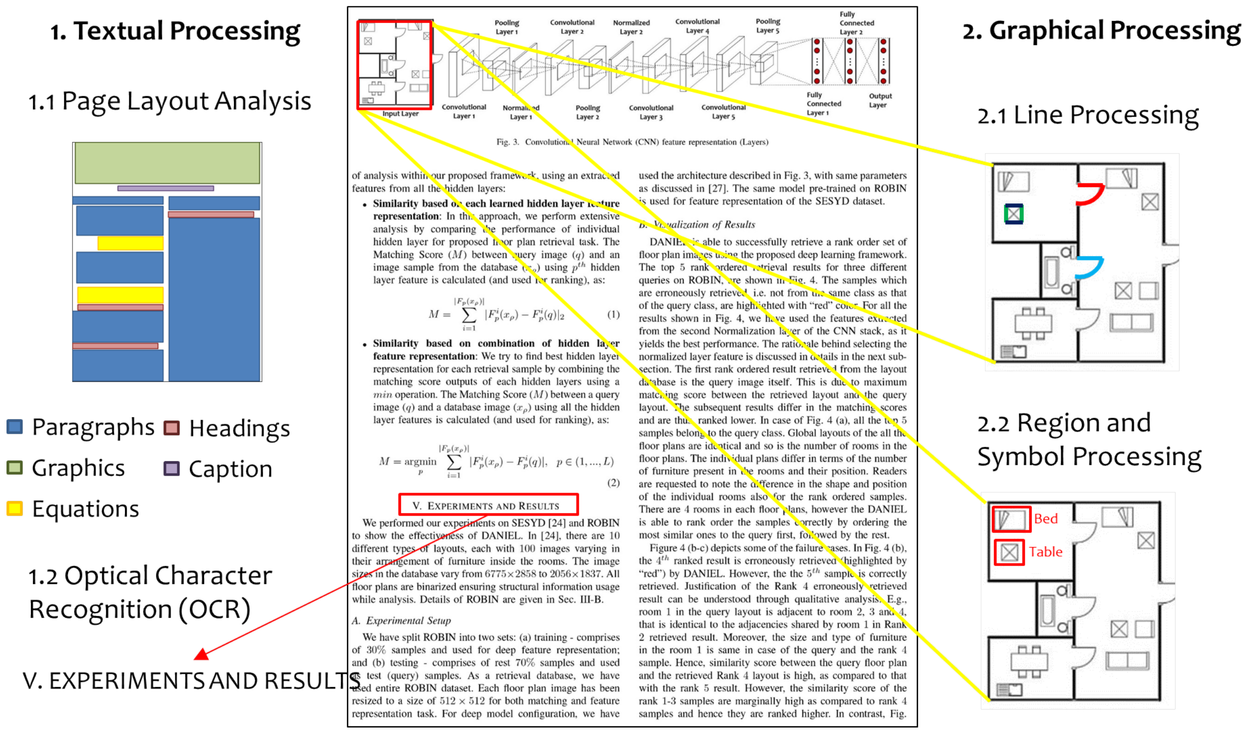


**Figure 1.1. :** Illustration of various subareas of document image analysis (DIA) on a sample scanned document [Sharma *et al.*, 2017].

different modes of query and assesses a wide variety of of content from the given query for retrieval purposes.

Design of a Content Based Image Retrieval (CBIR) system (refer Fig. 1.2) is a complex task and includes concepts from various academic disciplines such as image processing, computer vision, pattern recognition, machine learning, visual perception etc. Content-based means that the search analyzes the contents of the image rather than the metadata such as keywords, tags, or descriptions associated with the image. The term “content” in this context might refer to colors, shapes, textures, or any other information that can be derived from the image itself. CBIR is desirable because searches that rely purely on metadata are dependent on annotation quality and completeness. Having humans manually annotate images by entering keywords or metadata in a large database can be time consuming and may not capture the keywords desired to describe the image. The process of CBIR involves retrieving similar images in response to user’s search queries, using the inherent properties (features) extracted from the image itself. These features are mapped onto a feature space and the images with the most closely placed features in the feature space (depends on a certain metric) are deemed similar and retrieved from the database. An illustration of the same is shown in Fig. 1.2, where the context is chosen to be natural images. One more key aspect that plays an important role in a CBIR system is the modality of the query to the system. As illustrated in Fig. 1.3, query can be in the form of text, image, sketch, colour layout which depicts the colour combination that a user wants to look in the retrieved images or a concept layout where images that contain multiple related objects as specified in the query needs to be retrieved. However, the quality of the retrieval result of any CBIR system depends on the quality of content representation in the feature database. In a content based retrieval system, given a query and a database to search, the following questions need to be addressed:

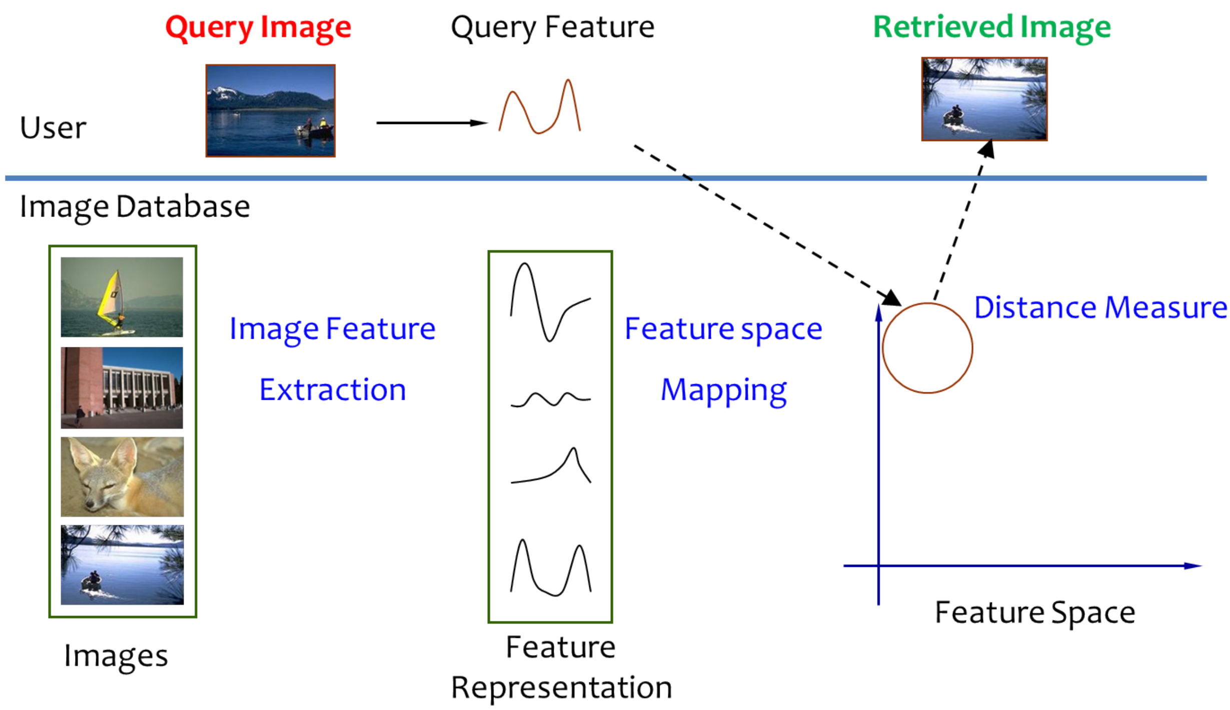


**Figure 1.2.** : An illustration of a basic Content Based Image Retrieval system.

1. What are the kinds of databases? This pertains to what kind of images are there in the database and what are the categories that the images have been classified into in the database.

2. What are the kind of queries? As illustrated in Fig. 1.3, the mode of the query can be an image, sketch, text, colour layout or a concept layout. Since each query mode has its own challenge, choice of modality is a critical research issue.

3. In which category does this image belong? What constitutes a category in the database and how one generates the feature set which represents each category present in the database distinctively.

4. In the given category, what are the "most similar" images in the database, with respect to the given query? Development of an efficient matching scheme, that helps evaluate the feature level similarity between the query as well as the categorized database images efficiently.

5. How does one make such searches efficient? What indexing mechanisms and efficient matching schemes can be used to make searching across databases fast and accurate.

Existing CBIR systems have been used for retrieval and have found applications in Art collections, Crime prevention, Geographical information, Remote sensing systems, Medical diagnosis, Retail catalogs, Face Finding, etc. Document image retrieval is a branch of CBIR that deals with extracting meaningful content out of document images such as, bank forms, cheques, scanned documents and then retrieving them for various applications. These days a great number of documents are scanned and archived in the form of digital images in digital libraries to make them available and accessible to the web. Therefore, a potential application of document image retrieval can be to make an online digital library where users would be able to do a search on a set of keywords, and get a list of relevant articles across the web, for viewing or printing. Architectural

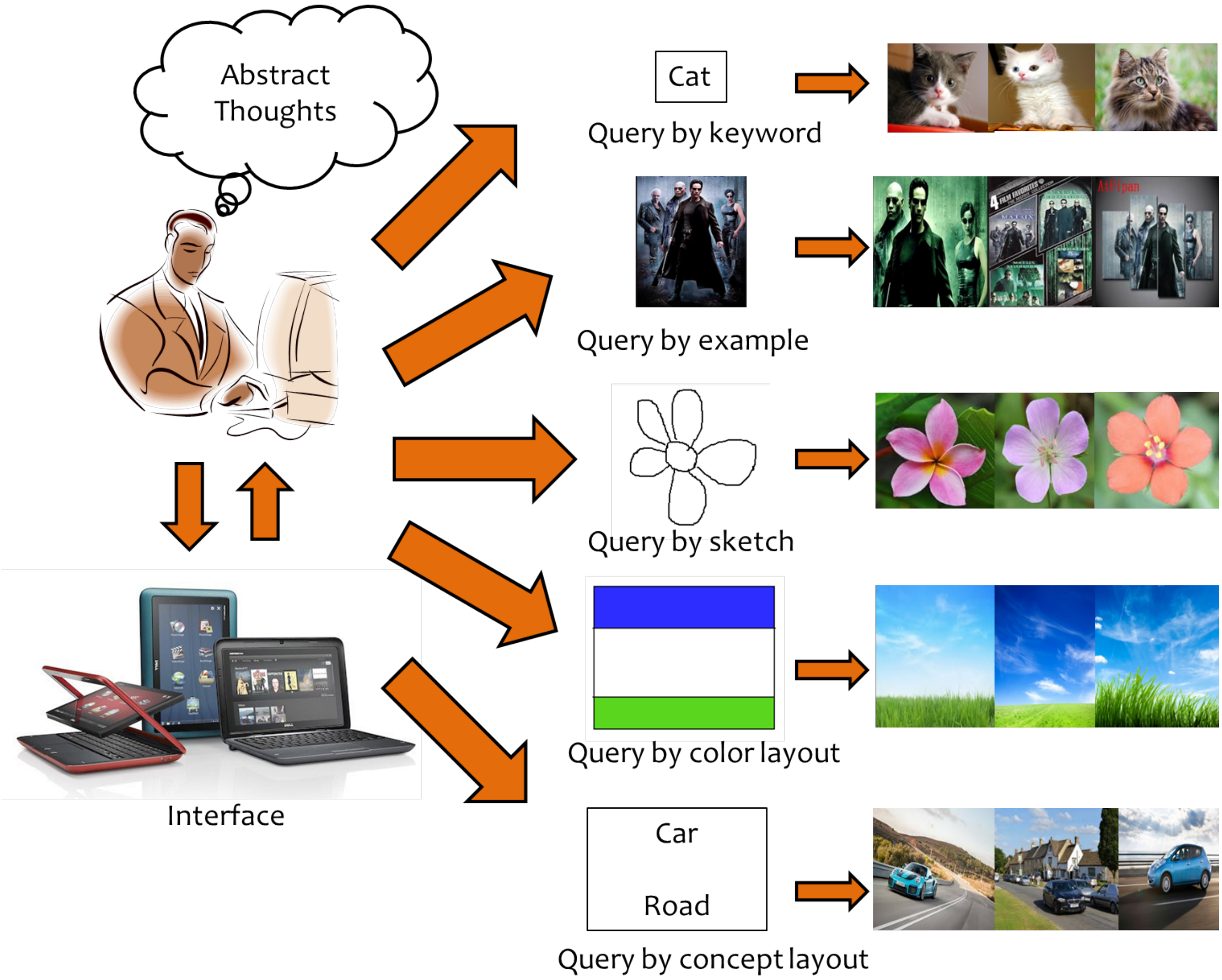


**Figure 1.3. :** Illustration of various query modes along with their corresponding retrieval results

floor plans are also a special type of documents, which have yet not been analysed with respect to its content and then retrieved based on similarity. Representing the inherent content present in a floor plan image/sketch automatically and identifying similar images from the database poses one of the fundamental challenge for the retrieval task. How to overcome the inherent challenges and come up with an efficient content based retrieval system specifically designed to represent content in architectural floor plans and retrieve thereafter, is a point which this thesis will reinforce.

## 1.1 MOTIVATION OF THE PROBLEM

The motivation behind the problem of retrieving similar floor plans by analysing meaningful content originated from the current needs and demands of the real-estate industry. To build a dream house in reality, customers visit architects, and specify the requirement "in words". An architect manually studies previous reference projects and suggests the best house and also searches for inspiration or a solution without having the final design at hand. This is an adhoc and tedious process. With the increase in number of online platforms, fast and automatic techniques for retrieval of similar projects are need of the hour. However, "*How to specify the requirements*

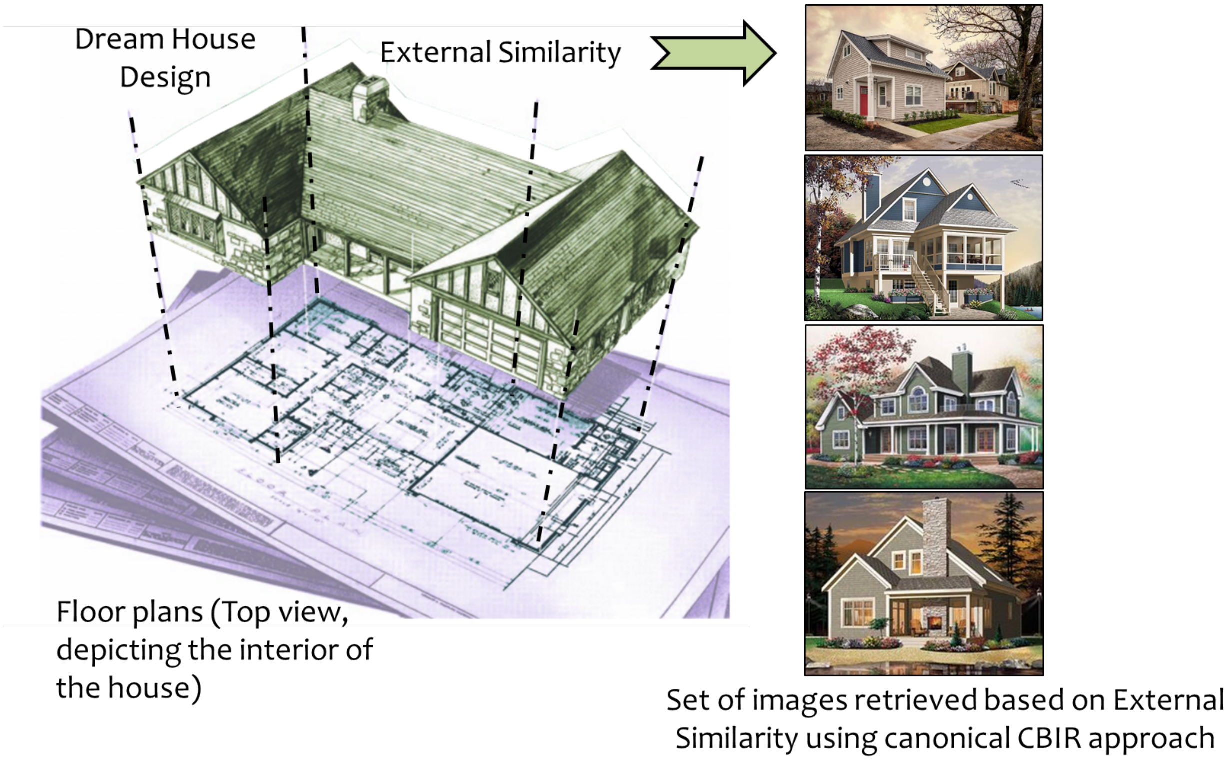


**Figure 1.4. :** Motivation of the problem.

*of a dream house?*" is still unclear. A naive solution is to follow the Content Based Image Retrieval (CBIR) approach to cater external similarity. Figure 1.4 shows such a scenario. It can be observed that the retrieved images of houses from the database are externally similar to that of the query image. However, there is no guarantee that these houses would match in their internal similarity as highlighted by the projection line and the floor plan.

However, the requirements of a user are usually pointing towards the internal structure of the home and the decors to be placed within each room. Consider a typical example where the requirement is specified as, "*the house must have* 4 *rooms, with* 2 *bedrooms with attached bathroom,* 1 *drawing room and* 1 *kitchen. The bedrooms must face towards west,...*", then interior similarities along with external structure similarities are also to be matched. Instead of using textual (subjective) query, a pictorial representation is always a better option. A floor plan is a fundamental form of architectural diagram, that gives a view of a house equivalent to a 2D horizontal cross section along with details of the interiors. Figure 1.5 depicts some examples of floor plan images corresponding to various types of houses.

Even though, floor plans are "*graphical language of communication*", however, investigation is required to answer "*how to match a pair of floor plans?*". In literature, researchers have carried out floor plan analysis in terms of symbol spotting and semantic segmentation. Symbol spotting in architectural floor plans deals with retrieving symbols from a database of floor plans, given a query symbol. Here the symbol set contains images of decor items like table, chair, bed, etc. It also finds the specific locations where the query image of the symbol is likely to be found. The other fields researched are floor plan analysis where existing works range from simple techniques for interpreting a hand-sketched floor plan for creating CAD models and automatically interpreting map drawings to analysing differentiation between thick, medium, and thin lines for segmenting

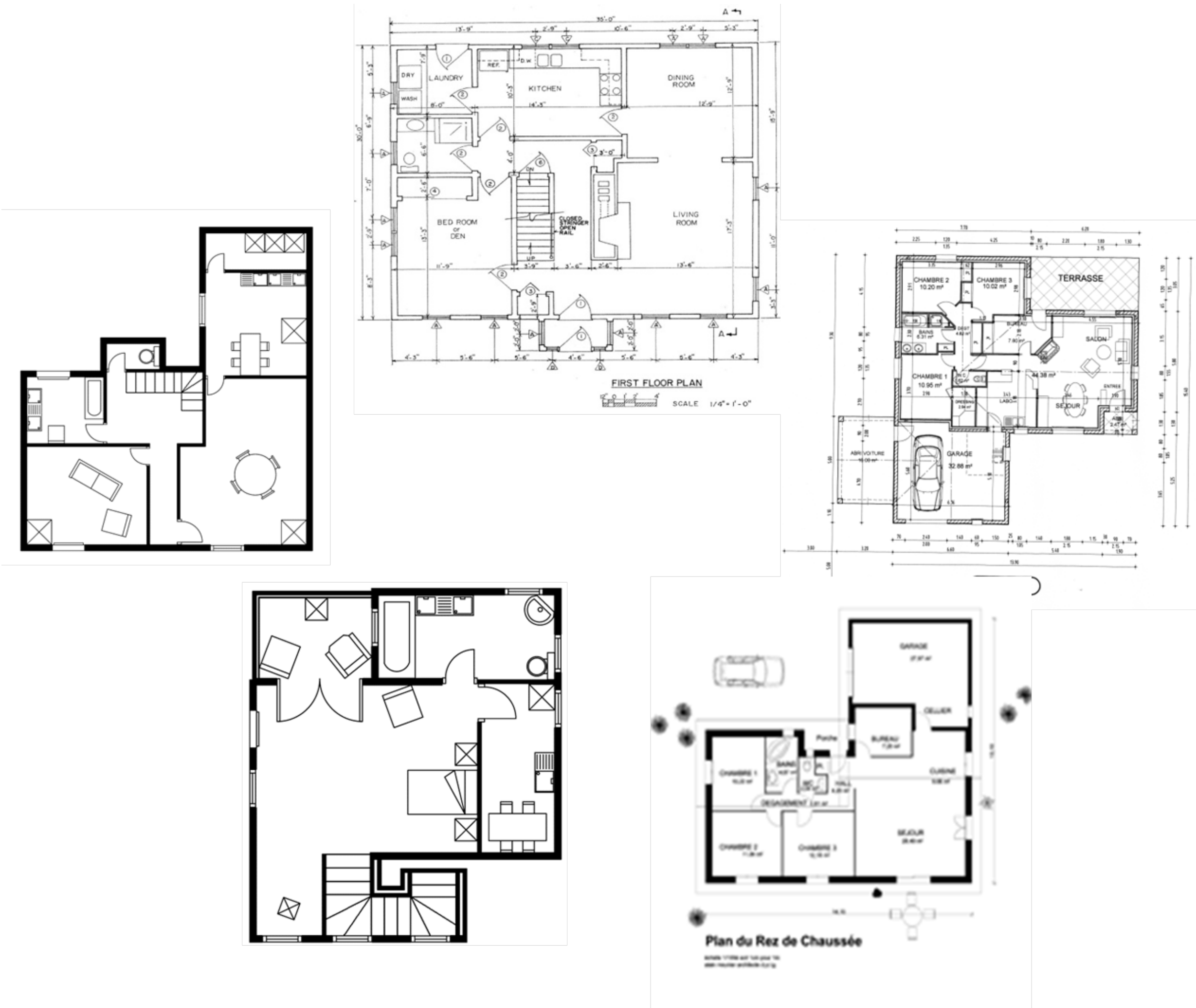


**Figure 1.5. :** Floor plans, showing 2D cross sections of buildings and displaying heterogeneity in design.

the room layouts in an efficient manner. Thus, upon doing a keen analysis of the existing work in the area of architectural floor plans, it was observed that retrieval in this particular domain is a challenging yet less researched area and has a lot of applications in today's digital scenario.

## 1.2 PROBLEM STATEMENT

The goal of the work proposed in this thesis is to design a content based retrieval system to perform the following task:

**Given a query image/sketch of architectural floor plans retrieve alike floor plans from the database, rank-ordered by similarity in content** (refer Fig. 1.6).

As proposed in this thesis, the input to the system would be a floor plan sketch or a floor plan image and the expected output would be similar floor plans as retrieved according to the specified content by the user. Content specification and appropriate feature set to represent the mentioned content plays a pivotal role in success of the retrieval system. Efficient techniques for


**Figure 1.6. :** Retrieval in floor plans by extracting meaningful content and looking for similarity across floor plan databases.

matching and retrieval are also proposed alongside capturing appropriate content from the floor plans.

### 1.2.1 Brief description of the Work Done

In Fig. 1.7, the problem of floor plan retrieval is represented alongwith its various possible solutions depicted in the form of a cube, containing a modality-axis, a feature-axis and an approach-axis. As shown in Fig. 1.7, the modality-axis of the cube corresponds to the type of modalities while querying. A content based retrieval system can have different modalities of representing the input query floor plan, for e.g. query can be an image (I) which represents floor plan images, a sketch (S), which can prove to be a convenient choice. Using pen and paper to scribble down ideas has always been a conventional way to query. The user looking for floor plans can easily draw an abstract floor plan with some interior details on a hand-held device and the sketched floor plan obtained thereafter can be used for querying. The third method can be a hybrid combination (H) of sketch and image representing a unified framework to capture the buyer's intent perfectly. Given a sketch or an image as query the framework should be able to retrieve similar floor plan images from the database. The term hybrid query mode is coined because in this mode the user can make online changes to the query, i.e. image or sketch, by adding or removing components in an existing floor plan. This mode currently serves as a future scope for this thesis. The feature-axis in Fig. 1.7 corresponds to the type of features extracted in the framework. The query and the existing layouts in the repository can be analysed for structural features as well as semantic features. Structural features correspond to the layout features that pertain to the room level analysis being used as content while extraction. These features use the topology and placement of rooms as matching criteria. Whereas, semantic features represent high-level features representing interior room details, sizes of the rooms, room-wise decor placement etc. It is to be noted that both of these features can be hand-crafted, which are typically

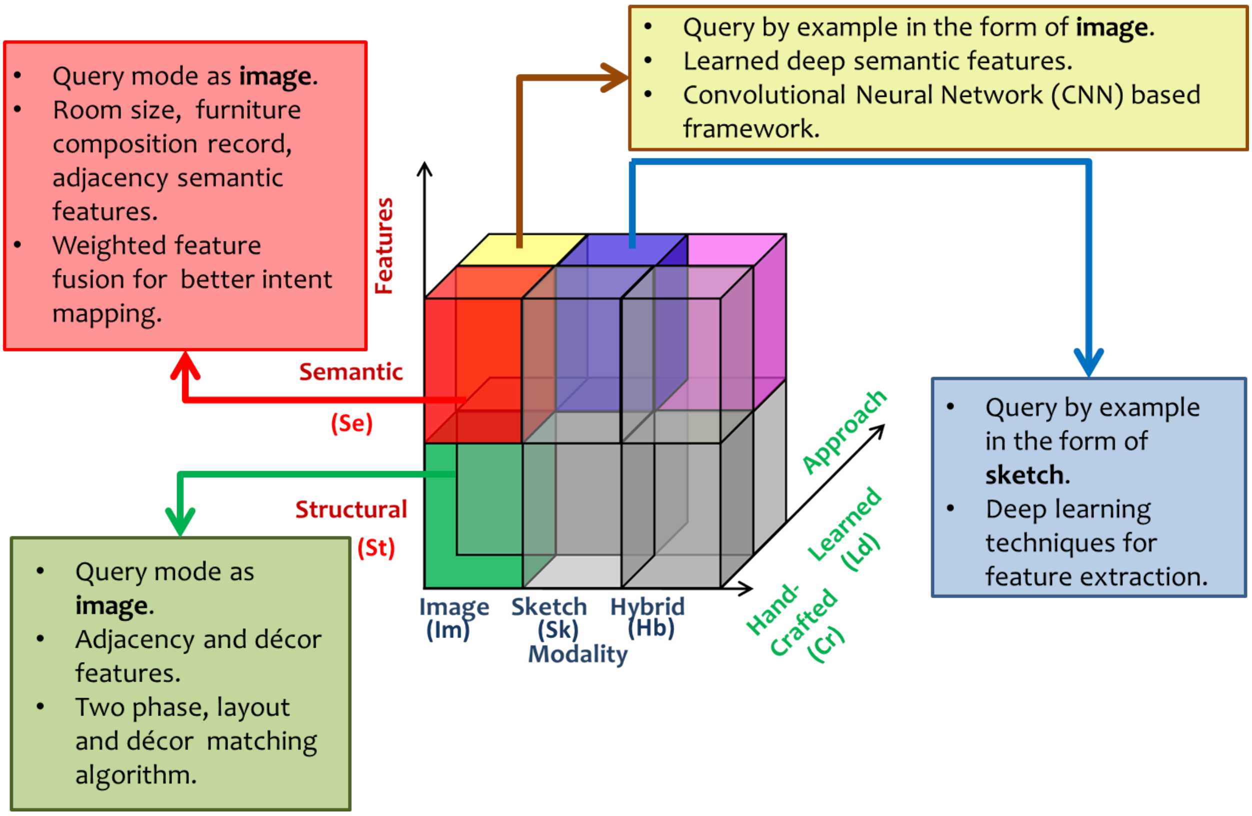


**Figure 1.7. :** Analysing floor plans for novel content aimed at retrieving similar layouts from a large repository.

chosen to represent the content in a floor plan, as well as, these features can be learned using a sophisticated learning framework, where explicit mentioning of the features is not necessary. The third axis of the cube in Fig. 1.7, represents such a scenario, where these two different approaches, i.e. the learned/ hand-crafted feature approaches are mapped. Thus, it can be inferred that such an amalgamation of various attributes while retrieval gives a great array of challenging problems to be dealt with while representation, analysis and finally matching and retrieval of floor plans. In this thesis, attempts were made to propose solutions to problems involving the following combinations (Im-St-Cr), (Im-Se-Ld), (Im-Se-Cr), (Sk-Se-Ld) as shown in Fig. 1.7.

## 1.3 RESEARCH ISSUES

Some of the key challenges encountered in the task of Content based retrieval and specifically floor plan analysis and retrieval are highlighted below:

1. In most cases, images/sketches require several stages of processing such as segmentation, smoothing, thresholding. The choice of an appropriate feature to correctly represent the content is in itself a big challenge.

2. As floor plans are 2D cross-sections of buildings, they are generally black and white and carry pixel information in a binary form. Moreover, there exists significant amount of whitespaces in the floor plan. Thus, throughout the image the information content is sparse in nature

which makes feature engineering a challenging task.

3. There is a dearth of publicly available datasets as architects refrain from sharing their designs. Thus, for implementation of an algorithm there is lack of publicly available sufficient data for the task of floor plan analysis and retrieval.

4. There is a huge-variation in the shapes and sizes of floor plans. This makes training a network, for instance, a deep learning framework, pretty challenging as resizing varied shaped floor plans leads to loss in information.

5. Sketch based query mode is in itself challenging to work with. Sketches incorporate a lot of user-variability and noise and require careful pre-processsing before analysis.

6. Representing image/sketch content using features demands a careful development of both hand-crafted and deep-learning features. The representation should ensure high intra-class similarity, at the same time high interclass dissimilarity. The representation should also be able to distinguish between overlapping classes. Moreover, the similarity criteria used for computing the match cost between the query and the model feature representation should also be unbiased and efficient.

***Assumptions***

The problem of content based retrieval in floor plans is solved under the following assumptions:

1. All the images are binarized.

2. Sketched drawings are scanned into image form for processing.

3. The floor plan images are digitally synthesised, i.e. they are not scanned and thus, free from any transformation like rotation, shear, scaling etc.

The focus of this thesis is on representation of floor plans and design of efficient similarity measures, under realistic situations. At the end of the thesis, an attempt has been made to bridge the gap between the two query modes, floor plan sketches and images using domain adaptation approach.

### 1.3.1 Organization of the Thesis

In this chapter, the importance of the problem of CBIR is introduced. Along with it a brief overview of the motivation behind a Content based retrieval for architectural floor plans is highlighted. The inherent challenges while proposing an efficient solution for this problem are listed. A brief outline of the approach to solve the problem is provided and the objectives that are trying to be achieved are listed.

In Chapter 2, a review of the existing work in the area of floor plan analysis is given, from which the motivation for this work is derived. Various aspects of architectural floor plan analysis dealing with symbol spotting, room level analysis and sketch based retrieval in floor plans are discussed. In the same Chapter, the datasets available in the floor plan analysis community are discussed. Also, basic image feature extraction strategies are reviewed. Some of the deep learning methodologies for feature extraction and domain adaptation are also discussed in Chapter 2, which serve as the motivation for the proposed multi-modal deep learning network.

Chapter 3 proposes a framework for the matching and retrieval of similar architectural floor plans under the query by example paradigm. Here query is taken as a floor plan image. A room layout segmentation and adjacent room detection algorithm is presented to represent layouts as an undirected graph. Also a novel graph spectral embedding feature is proposed to uniquely represent the layout of the architectural floor plan. This graphical approach helps in effective and efficient matching of the room layouts. Room semantics in terms of both the room structures and room decor is used to retrieve similar floor plans from the repository. To match the semantic similarity between a pair of floor plans, a two stage matching technique is proposed and high retrieval accuracy is obtained.

In Chapter 4, a deep learning paradigm to extract both low and high level semantic features from a layout image is proposed. The key contributions in the proposed approach are, a novel deep learning framework to retrieve similar floor plan layouts from repository and analysing the effect of individual deep convolutional neural network layers for floor plan retrieval task.

In Chapter 5, a novel end-to-end framework for extracting high level semantic features like area and room-wise decor arrangement for the task of fine grained retrieval is proposed. Further, a technique to perform feature fusion to aggregate high-level semantic features extracted is also proposed. Weighted feature fusion helps in setting preferences to particular features while retrieval and satisfying specific user demands.

In Chapter 6, a novel system which uses sketch based query mode to retrieve corresponding similar floor plans from the repository is proposed. Two approaches are discussed, one with a system using Cyclic Generative Adversarial Networks (Cyclic GAN) for mapping between sketch and image domain. The key contributions here being a novel sketch based floor plan retrieval framework using an intuitive and convenient sketch query mode and conjunction of Cyclic GANs and Convolution Neural Networks (CNNs) for the task of hand-drawn floor plan image retrieval. The second approach being the usage of autoencoders along with Cyclic GANs for improving the performance of the retrieval framework and helping in better domain mapping between sketch and image domains.

Finally, in Chapter 8, the contributions of this thesis in the field of content based retrieval in architectural floor plans are summarized and a few possible extensions and future prospects for the work are identified.

Two appendices are also provided at the end of this thesis, where, Appendix 1 details out the creation of the floor plan datasets, Repository Of BuildIng floor plaNs (ROBIN) and Sketched-Repository Of BuildIng floor plaNs (S-ROBIN) that have been specifically designed to aid the analysis and retrieval in floor plans. Appendix 2 contains the description of a graphical user interface, designed for the task of interactive floor plan retrieval.