In recent years, digital cameras have become ubiquitous; storage is less expensive, Internet access is available nearly everywhere and digital social interaction is an increasingly popular trend. Due to these reasons, digital images have grown exponentially and have been making it beyond the abilities of people to easily manage these important contents. In an effort to solve this burden, the author investigates on image understanding in order to bridge the semantic gap between human and machine. Towards this goal, the author proposed image analysis methods and system designs that go beyond the superficial image content analysis. The proposed schemes (i) fully exploit the holistic content analysis by utilizing not only the whole original image, but also its salient regions and its background; (ii) leverage other related information about the image such as GPS, temporal, layout, optical, and contextual information; or (iii) combine these schemes to complete this difficult task. The author also examines user’s behaviour, user’s perception, aesthetic values and photography grammar.

In the scope of this dissertation, the author focuses on automatic image annotation, result re-ranking, and categorization and quality assessment tasks. These tasks are among the most fundamental and essential ones for semantic understanding of image. The contents of the thesis can be summarized as the following.

Chapter 1 sets the stage by giving the background of the research problem as well as the scope of the thesis namely, automatic image annotation, result re-ranking, and categorization and aesthetics quality assessment.

Chapter 2 gives the state-of-the-art research work on the related techniques towards image understanding, and the positioning and contributions of this thesis in this regard.

Chapter 3 explores the problem of automatic image annotation in a general case. One of the main bottlenecks in this area is the lack of integrity and diversity of features. The author proposes to solve this problem by utilizing 43 image features that cover the holistic content of the image from global to subject, background and scene. In the approach, salient regions and the background are separated without prior knowledge. Each of them together with the whole image are treated independently for feature extraction. Extensive experiments were designed to show the efficiency and the effectiveness of the approach. Two publicly available datasets manually annotated with the diverse nature of images were chosen for the experiments, namely the Corel5K and ESP Game datasets. The results confirm the superior performance of the proposed approach over the use of a single whole image using sign test with p-value < 0.05. Furthermore, the proposed combined feature set gives satisfactory performance compared to recently proposed approaches especially in terms of generalization even with just a simple combination. The approach also achieves a better performance with the same feature set versus the grid-based approach. More importantly, when using the proposed set of features with the state-of-the-art technique, the results show higher performance in a variety of standard metrics.

Chapter 4 focuses on the problem of automatic annotation in the personal case. By analysing users’ behaviour and technology trends, the author proposes a novel solution for this task. The method integrates all contextual information available to and from the users, such as their daily emails, schedules, chat archives, web browsing histories, documents, online news, Wikipedia data, and so forth. Subsequently, the integrated information is analysed and important semantic terms are extracted. The keywords are in the form of named entities, such as names of people, organizations, locations, and date/time as well as high frequency terms. They serve as annotation candidates for the photograph. Users can choose to validate these candidates. Experiments conducted with 10 subjects and a total of 313 photos prove that the proposed approach can significantly help users with the annotation process. The approach achieves a 33% gain in annotation time as compared to manual annotation. The results also demonstrate encouraging accuracy rate of the suggested keywords.

Chapter 5 is dealing with results re-ranking in the image retrieval task. Image search systems have a very limited usefulness since it is still difficult to provide different users with what they are searching for. This is because most research efforts to date have only been concentrating on relevancy rather than diversity which is also a quite important factor, given that the search engine
knows nothing about the user’s context. In the chapter, the author describes the proposed approach for photographic retrieval task (within the scope of ImageCLEF 2008). The novelty of the approach is the use of AnalogySpace, the reasoning technique over commonsense knowledge for document and query expansion, which aims to increase the diversity of the results. The proposed technique combines AnalogySpace mapping with other two mappings namely, location and full-text. Re-ranking mechanism is employed to the resulting images from the mapping by trying to eliminate duplicate and near duplicate results in the top 20. The experiments and the results conducted using the IAPR TC-12 photographic collection, with 20,000 still natural photographs, are represented. The results show that the integrated method with AnalogySpace yields better performance in terms of cluster recall and the number of relevant photographs retrieved by maintaining precision. The author finally identifies the weakness in the approach and ways on how the system could be optimized and improved.

Chapter 6 is interested in the problem of high quality photo categorization and aesthetic quality assessment. The chapter outlines the proposed framework for the tasks. The author addresses these challenges by exploring the aesthetics from the combined perspectives of the artists and photographers. The author proposes to use the aesthetic primitives of images for visualization as a guideline for high and low-level image feature extraction and to classify this high quality content into six creative exposure themes, which are commonly followed by the professional photographers. Furthermore, the proposed framework suggests evaluating the quality of the photograph accordingly to these themes. In the proposed approach, the tasks are solved using statistical modelling and learning schemes. A small experiment using only the camera setting features was conducted and the result was encouraging.

Chapter 7 concludes the findings. Then, the future perspectives in structuring the image collections and eventually in making sense out of them are presented.

These analysis and methodology designs presented in the thesis shall contribute to the better understanding of visual content beyond the conventional approaches. In addition, it is shown that they meet one or more of the user’s requirement attributes. Therefore, many fully targeted visual related applications and services - not limited to the image related ones - could rise from these findings.