PhD Probation Report

Adam Rae

27th August 2008

Contents

Chapter 1

Introduction

Around 16,000 years, a community of humans living near caves at what is today known as Grotte de Lascaux, France daubed primitive paints on the wall to show scene of hunting and the world around them. They left representations of the world for generations of people to see and share. It is through the markings like these that we can infer a little more about life millennia ago, and they demonstrate the importance of images to our collective knowledge.

While art and graphic design developed over the intervening years, it wasn't until the 14th century that audio was recorded in the form of rotating cylinders used in mechanical bell-ringing systems and clocks. It took until 1830 for the first moving pictures to be recorded as a string of still images that were played back in quick succession to give the impression of movement. With the invention of electronics and digital electronics in particular these various forms of media really began to take off as they could be captured and transmitted around the world.

> Nation shall speak peace unto nation - British Broadcasting Corporation motto
>
> While this may have been a noble sentiment to aspire to in the 1920s when this motto was first coined, I believe technology has surpassed the ideas behind these words to enable individuals to speak to other individuals without regards for national borders in many new ways, thanks to multimedia technology.

As the cost of production, storage and dissemination has plummeted for images, audio and video, challenges have arisen regarding how to most effectively handle this wealth of information. Digital recording devices have become cheaper, more widely available and are able to recording in better quality than ever before. Media representation has given us lossy and lossless file formats suitable for many different situations, from satellite broadcast to mobile streaming, each with their unique requirements and capabilities. Digital file storage devices have grown more capacious, high performing and more reliable.

The one area that has not caught up is *how to search and retrieval multimedia* efficiently and effectively in these new, vast data sets and this is the focus for my work. More specifically, I wish to investigate the relationship between social context data and its effect on multimedia information retrieval.

Chapter 2

Related work

2.1 Introduction

To highlight the the current weaknesses in existing approaches and to emphasize useful models and methodologies, this brief literature review is split into four main areas which will lead into conclusions based on existing evidence in the final section.

This project is focused on how best to use various feature descriptors within the domain of image-based information retrieval when combined with data provided by online social networks. It is this context that has driven the literature review strategy used within this report.

The criteria for reviewing the literature include its significance to my particular area of interest, and this can be in terms of conclusions, the methodologies it uses to gain those conclusions and even its review of established work which is also relevant to my own. I have considered how influential a work is in terms of its prevalence within the community, but not to the exclusion of work that has yet to be published widely if I have considered it to be important.

2.2 Multimedia search - existing systems and state of the art

Extensive research has been carried out in the field of text retrieval and underpins large scale systems like Google, Yahoo! and MSN that handle predominantly text-based web pages. Compared to other media forms, text is easy to store and computationally less expensive to transform and analyse. This means that text-based systems are more scalable as a result - a point which is important when considering large-scale Internet based systems.

Due to this existing body of work, many existing multimedia search systems focus on the text metadata associated with a piece of media and treat a datum as nothing more than an attribute of the metadata (de Jong, Westerveld, and de Vries 2007). While understandable, this approach limits itself by excluding all the information contained within the piece of media and ultimately contributes to the inability of existing systems to fully satisfy the information need of users (Conole 2001). The area of text based information retrieval is without the scope of this project and so related literature is only superficially discussed.

New techniques in the handling of images, video and audio have yielded systems capable of using the information from the media stream itself. Images can be automatically categorised based on their visual features, video can be segmented and repackaged based on what story it is telling (Heesch, Pickering, Rüger, and Yavlinsky) and audio can be fingerprinted and compared to millions of other songs in a few seconds (Cano, Batlle, Kalker, and Haitsma 2005; Haitsma and Kalker 2003).

These new techniques can only be used once the user has expressed their query and presented it to a system. Different input methods have been investigated (Li, Xie, Liu, Tang, Li, and Ma 2004) and tend to fall into one of two categories; query-by-description (QBD) and query-by-example (QBE). The first uses the established body of research on text-based metadata search to take a keyword or description based query and use this to search a data set. The second, invites the user to provide an example image, video or sound clip and to use this as a starting point for the search engine to use. The example can be in the form of an existing piece of media that is otherwise available and similar enough to the requirements of the user, or the user could be asked to create an example at the time of entry, through a visual sketch or whistle a tune. It is this QBE approach that this project is primarily concerned with, but not to the exclusion of useful QBD techniques and knowledge.

Once the information need of the user has been expressed it must be transformed into a representation that the system can use to compare to an index of the others it already contains. This is normally done by extracting visual features from an image based on colour, texture and shape which are then combined to form vectors that characterise a particular image. These feature vectors are far smaller than the original piece of media and can therefore be more quickly compared. The MPEG-7 specification and its satellite papers (Manjunath 2001) contain very useful overviews of some prominent features and their calculation.

The choice of features for a system is very important and depends on the context of the system and the data it manages. A system designed for one form of search task (X-Ray images from a hospital radiography group for example) may have different general trends in the type of features they find most useful and that yield the most relevant results when compared to another. Even within a search task context, different users may find different features to be useful.

Deriving a user's profile to take advantage of their feature preferences is another important strand to this project's research.

This profile that characterises a user's preferences is likely to be both dynamic and noisy. The former, due to the way people's interactions with a system change over time, and the latter due to the inability of a system to fully interpret a user's information need, and so it is possible that data derived from a user's interactions could be inaccurate reflection of what the user had in mind.

The eventual combination of evidence from multiple features is also important. Once a set of preferences for particular features have been learnt, the results from each 'expert' must be combined. This multi-modal evidence issue crosses disciplines and much of the work carried out in IR has had roots in areas like political science with voting systems that combine ranked preference lists of political candidates - for example the Borda Count method. This method is both simple to implement and higher performing than rudimentary round robin algorithms. There is criticism of this approach in the literature, but only from the perspective of its use in multi-agent systems, where values used in the method can be misrepresented to manipulate the system to the advantage of a particular agent. This is not an issue in IR as all hypotheses are combined in a system that itself defines the parameters to use and is not reliant on external, possible faulty data.

2.3 The utility and weaknesses of Content Based Image Retrieval

The three main categories of features used to represent the visual content of an image are colour, texture and shape (or form). These can be evaluated on the criteria of repeatability, distinctiveness and robustness. A descriptor must be able to distinguish between visually similar images and yet be robust to noise (be it deformation, or through information corruption or loss). It must also produce the same description for the same image each time it is applied so that when feature vectors from multiple images are compared, an accurate and robust comparison can be made. This last criteria is vital for the production of indexes which form the backbone of most search engines.

Prominent feature analyses Bay *et al.* (Bay, Tuytelaars, and Gool 2006) tackle the issue of feature descriptor robustness directly in their evaluation of their Speeded Up Robust Features detector/descriptor scheme based on the Hessian matrix. Their focus on scale and rotation invariance lead to a system that was both fast and outperformed GLOH, SIFT (see below) and PCA-SHIFT algorithms considered state-of-the-art. This work is important in that it took existing methods and was able to improve performance considerable by selectively reducing complexity in both the detector (by using approximations and integral images) and the descriptor without sacrificing distinctiveness. This trade-off shows that this type of compromise can be made in order to ultimately improve such a system.

The work of Lowe (Lowe 2004) focuses on the issue of feature distinctiveness by building up a database of recognised objects and matching items in scenes to that prior knowledge, partly involving the Scale Invariant Feature Transform (SIFT) method. By being invariant to certain transformations the system is able to discriminate between images of different artifacts and images of different views of the same or similar artifacts. It does this by producing high-dimensional vectors representing image gradients in local regions of the image. Its distinctiveness comes about by being able to match high dimensional key-points to noisy or blurry sections of images, whilst also being able to match low-dimensional key-points to unoccluded parts of objects. It uses a cascade filtering approach whereby features that are computationally expense to compute are avoided in favour of cheaper ones until absolutely necessary. I find this approach to be particular interesting in that the performance gains yielded by such a method would be particular useful when used with very large scale datasets such as Flickr. The problem is that for this particular implementation found in the paper, repeatability diminishes as the dataset size increases.

Mikolajczyk and Schmid (Mikolajczyk and Schmid 2005) extend the work of Lowe by first comparing different methods for image matching and object/scene recognition based on computing local interest regions, including SIFT. The criteria of precision[1] and recall[2], the two primary metrics of information retrieval, lead them to show that of those methods analysed, certain descriptors worked best for certain scenes (shape context performed well except for scene with edges that were difficult to detect) but overall Gradient, Location and Orientation Histogram (GLOH) and SIFT performed best. Their ultimate conclusion that robust region-based descriptors performed better than point-wise descriptors is an important point to take into account when selecting features to be extracted from the Flickr dataset.

Colour based feature descriptors have been shown to be high performing and relatively easy to compute (Feng, Zhou, Shen, and Pan ; Ionescu and Ralescu ; der Weken, Witte, Nachtegael, Schulte, and Kerre 2005) as well as usually being tolerant of image transformation and as such, are useful for image dataset indexing. They have been shown to be particularly useful in task specific matching, like face and human skin-tone detection (Chai, Phung, and Bouzerdoum 2003; Singh, Chauhan, Vatsa, and Singh 2003; Vezhnevets, Sazonov, and Andreeva 2003).

As an overview comparison of texture based features, the work of Howarth and Rüger (Howarth and Rüger) is particularly useful as it was focused on their performance within the context of query-by-example image retrieval as opposed to the more usual classification task, as well as using the whole image as opposed to local regions. Its comprehensive evaluation of grey level CD-occurrence matrices, Tamura and Gabor techniques with respect to these contexts was novel. What is also important is that performance was always improved by introducing colour descriptor information to the texture based features. This multi-modal combination or data fusion technique is particularly important and the work of Bartell *et al.* (Bartell, Cottrell, and Belew 1994) in combining evidences exemplifies this.

Distance Measurement Once features have been extracted from a set of images, their proximity to each other needs to be calculated in order to provide quantative values for the similarity for each image to each image. Those systems that are designed around static datasets tend to calculate these relative distances before users can access the system to speed up performance[3], but whenever the dataset alters, the distances and ranks must be recomputed and added to the database.

The choice of distance metric is important as there are different ways of calculating proximity, from relatively simple Euclidean, correlative and covariance measures (Stevens and Beveridge 2000), to more complex Hausdorff distance maps (Baudrier, Millon, Nicolier, and Ruan)and Kantorovich values (Kaijser 1998). These more computationally complex distances are not always consistently useful, but can be used when other methods fail. The work of Hu *et al.* (Hu, Rüger, Song, Liu, and Huang) gives a good experimental analysis of different methods and shows that even simple methods can be high performing

[1] Precision is the fraction of documents returned by a search engine that are judged to be relevant to the user's query.

[2] Recall is the fraction of relevant documents that are retrieved out of the total number of relevant documents.

[3] üBase is one example based on computing ordered lists of similarity - http://kmi.open.ac.uk/technologies/ubase/

(Cityblock for example).

Some methods have been developed for grey scale images but then transformed and tested with colour images as well (der Weken, Witte, Nachtegael, Schulte, and Kerre 2005).

The work of Jia and Kitchen (Jia and Kitchen 2000) focused on more directly measuring image similarity by not differentiating the feature extraction and comparison stages and was particularly interesting for the use of an inductive learning algorithm (C4.5) to predict the class of object found in a scene. This was also shown to be invariant to scale, rotational and translational transformation. In a similar vein, the work of Arnia *et al.* (Arnia, Iizuka, Fujiyoshi, and Kiya 2007) take advantage of coefficient calculated during JPEG compression to provide discriminating data, again bypassing the feature extraction stage providing a method that is both effective and cheap to compute in appropriate situations.

There are two main weaknesses of CBIR; the first is demonstrated in this Section - there a many different features that can be extracted from an image, which itself may be represented in many different formats, as well as many different ways of computing distance and combing multiple evidences. The choice of which of these components to use and in what processes differs between systems and contexts (and opinions) (Deselaers, Keysers, and Ney 2004). This choice is vital for the performance and effectiveness of a system and while some combinations of components have been shown to work work in particular scenarios, there is not yet a master mapping between 'system context' → 'appropriate systems components'.

The second is what is termed the 'semantic gap' (Enser and Sandom). This describes the difference between descriptions of the same data by different linguistic representations - for example a high-level written description in English compared to a low-level computed feature description vector. Both describe the data, but in different ways and from different perspectives. This has big implications for information retrieval, as queries are often entered into a search system in one representation but the system has to interpret this in order to use the representation it uses to index data.

2.4 The potential of social networks in providing more useful data

Networks and graph theory in general can trace their roots back to the work of mathematicians like Euler with his Königsberg Bridges problem. Cycles on polyhedra were later studied by Kirkman and Hamilton in the mid 19th century, which lead to the concept of the Hamiltonian graph (one which contains a Hamiltonian Cycle, or path that visit each node in a graph exactly once). The algorithms used to traverse and analyse these graphs become a major part of discrete mathematics research. This work continued to develop, but it was only when scientists began to obtain large amounts of interconnected data that the field of large scale network analysis really began to flourish. Names like Albert-László Barabási, Duncan Watts and Steven Strogatz have introduced the particularly important concepts of scale-free networks (those networks, similar to many “real world” networks, whose structure and dynamics are independent of the size of the system) and the small-world phenomenon (where even in very large networks, the average inter-nodal path length is relatively small - see the famous Six Degrees experiment of Stanley Milgram, 1967).

In his book *Linked: How Everything is Connected to Everything Else*, Barabási does not merely highlight the existence of networks and particular scale-free networks in everyday life, from molecular interactions to the Internet, but stresses why this knowledge is important. By knowing, for example, how to find nodes that act as hubs within a network, the network can be made more resilient to attack or degradation. This structural knowledge can also show how to best exploit the network by predicting how it will continue to grow and the effects of structure-changing events.

This has particular relevance within the context of multimedia search as the derivation of clusters within a network (Ahn, Han, Kwak, Moon, and Jeong 2007; Aida, Ishibashi, Takano, Miwa, Muranaka, and Miura 2005) can be very useful in informing recommender systems. These systems occur throughout the web, including media sharing websites like Facebook[4], Flickr[5], Picasa Web Albums[6] and Bebo[7]. These systems can make a range of recommendations from simply reminding a user that another user in their social network has uploaded new images and bringing this to their attention, to specifically showing a user images relevant to their social context. Flickr, as a site based on media sharing, provides ways for a user to explore photos based on their ‘interestingness’ as defined by their algorithm that takes into account the visual content of the image.

Although users form explicit groups in online photo sharing websites by affiliating themselves with other users who share similar attributes (like location), tastes (like black and white photography) or behaviour (uploads from mobile phones), clusters may also form that describe implicit groupings. These sets of users who are in close proximity to each other in terms of their characteristics may be unknown to the users themselves but can be exploited by a search system to make recommendation or to influence search results ranking by taking into account the preferences of users within the implicit cluster. Currently, systems like Flickr don’t make use of these implicit clusters.

2.5 Social networks

My work will be focused on data gathered from the Flickr online photo sharing website. This is due to having a supervisor who is based at Yahoo! Research Barcelona who has expertise in analysing the system and because it provides me with a data set and test bed that will be particularly useful for my research. There are millions of users and billions of images and videos, most of which have information externally available through both the website and the public API. Users can define relationships between themselves by stating that another user is a ‘Contact’, ‘Friend’ or ‘Family’ member, although there is no restriction on how these labels are applied or how many relationships can be specified. There

[4]http://www.facebook.com
[5]http://www.flickr.com
[6]http://picasaweb.google.com
[7]http://www.bebo.com

is a broad spectrum of user interaction with Flickr, from some users who only upload a handful of images to share with their immediate family, to others who upload thousands of images and have thousands of 'Contacts'. As this particular relationship is explicit, easily understood and the data is easily acquired (it takes only one API call compared to at least three to derive information for a 'Comment' relationship based on one image), it tends to be the focus of work based on understanding the structure of the Flickr social system. As it is a one way attribution ("I declare this person to be my contact") and does not have to be mutually agreed, this relationship forms a directed graph, unlike the 'group affiliation' relationship which is undirected.

The 'group affiliation' allows users to affiliate themselves with others in groups of shared interest or attribute (e.g. users from the same town). Users are also able to comment on and modify the metadata of other people's media. All these interactions build up to produce not just one, but many social networks based on the same users and their media.

Due to commercial sensitivity, very few statistics regarding Flickr are publicly released by the company. Very little work has been carried out by the academic community to analyse this particularly rich multimedia and social network dataset. The most extensive external analysis is that of Negoescu (Negoescu) based on analysing directed graphs of 3,544 Flickr nodes, which provides strong evidence to support the idea that the Flickr social network of 'Contacts' exhibits small-world characteristics, with a mean inter-nodal distance of $D_{av} = 3.88$ (so in this case, Flickr could be said to be a world with 3.88 degrees of separation) and a clustering co-efficient of $\bar{C} = 0.13212$. The clustering co-efficient quantifies how far a vertex and its neighbours are from being a complete graph[8] and a value relatively higher than that of a randomly generated graph indicates that the structure of the graph is less homogeneous and tends towards clustering. It is calculated in the following stages:

$C_i = \frac{|\{e_{jk}\}|}{k_i(k_i-1)} : v_j, v_k \in N_i, e_{jk} \in E,$

where C_i is the clustering coefficient for a node in a directed graph $G = (N, E)$ where N is a set of nodes and E is a set of edges. The edge e_{ij} describes the edge between vertex v_i and v_j. Edges are directed, therefore $e_{ij} \neq e_{ji}$. Each vertex v_i has k_i neighbours given by the sum of edges incident to and departing from the node, which is based on the definition of a neighbourhood:

$N_i = \{v_j\} : e_{ij} \vee e_{ji} \in E$ where $k_i = |N_i|$.

The overall clustering coefficient for a network is then given by the graph-wide average

$\bar{C} = \frac{1}{n}\sum_{i=1}^{n} C_i.$

where n is the total number of vertices in the graph.

This coefficient value in the work of Negoescu was shown to be two orders of magnitude higher than those of randomly generated graphs based on the same vertex set and therefore, according to the definition given by Watts and Strogatz (Watts and Strogatz 1998), can be considered a small-world network.

Similar work was carried out by Ortega and Aguillo (Ortega and Aguillo 2008), which focused on the nature of groups on Flickr, but they also calculated

[8] A COMPLETE GRAPH is an undirected graph such that for every pair of vertices that make up the graph there is an edge between them so that every vertex is connected to every other vertex.

values for the clustering coefficient for their dataset. For their set of 663 nodes, $\bar{C} = 0.282$. While this is a different value to that of Negoescu, it is of the same order and is still far larger than that of a random network developed with the same nodes (where $\bar{C} = 0.005$). The difference could be explained by the dataset selection process, resulting in sets one of which is 5 times larger than the other that may more accurately reflect the Flickr dataset as a whole.

While the numbers emerging from this body of work are very useful at setting the scene, it must be remembered that these values are indicative and not comprehensively representative. The data set used in this work was gathered in a pseudo-random manner (random access being impossible using the publicly accessible mechanisms) which while unavoidable does mean that questions of representativeness arise (this also might go some way towards explaining the differences between the values of Negoescu and Ortega (Ortega and Aguillo 2008)). This set was then pruned to leave one 'giant component' of users who shared at least one connection with another user in the same set of users, essentially removing all those users who were orphaned by the selection process who or didn't connect in any way to other users on Flickr. This means that results generated on this set are not representative of the Flickr system as a whole.

Even though there are flaws in the process, the evidence currently suggests that Flickr is small-world due to the coefficient and mean distances results of both the work of Negoescu and Ortega.

Another interesting point highlighted by Negoescu was the definition of the 'Contact' relationship in the Flickr system represented by a directed edge in the graphs analysed in his work. He suggests that due to the existence of users with exceptionally high numbers of 'Contacts' and other with none, this label was used not in the sense of a real world friend or colleague but more as a form of book mark, or link to a person and their presence in the Flickr system.

This leads to an interesting further question:

> How do the definitions used by a community arise and how do they translate into online representations?

Although it addresses the area of Digital Rights Management, the article from the BBC (News 2008) highlights the way communities transfer real world relationships, and their associated mores, into an online environment. In this case Australian Aboriginal communities have transferred their real life relationships with their family, friend, their own community and other communities and the behaviours associated with each kind of relationship into an environment of shared archival media. The system that developed around this real world network meant that in line with tradition males weren't shown images of female rituals and rituals of other communities could only be shown if permission had been explicitly obtained.

This same process of community driven definition happens whenever users start to interact online. This is how folksonomies grow to reflect the needs and usage of users when producing tags for example. Different systems encompass this interaction in different ways. Flickr explicitly defines a few labels (like 'Contact', 'Friend' and 'Family') and it is up to users to use them as they will. Others (e.g. Facebook) have a far larger range of more complex labels that allow a user to more accurately describe their relationship with another user. They can also define no explicit relationship with a user, merely having an unnamed 'connection'. These different types of relationships form strata in the overall social network of a system, and little work has been carried out to find out how these strata interact and how they can be useful when used together.

Search by using real world social networks has been demonstrated before (Granovetter 2008; Dodds, Muhamad, and Watts 2003), but it is only recently that work has been carried out to see how this paradigm transfers online (Watts, Dodds, and Newman). Watts showed how the architecture of many systems that existing on the Internet are comparable to real world systems and so the methods and techniques used there can be transferred and how social networks can allow efficient decentralised search with minimal data.

In principle, social networks give more contextual data about a user. With more data, a system can interpret the information need of the user more accurately and provide better search results in terms of performance, relevance and interest. By understanding the trends within a single user's behaviour profile as well as those found in the interactions between users, they can be exploited to provide a more personalised search system. But in order for a system to become better, it must also gather information as to how well it currently performs and in what way, in order to improve. For this, it requires feedback from the user.

2.6 The importance of feedback

Feedback in the area of information retrieval has been a major theme in recent work. Relevance feedback (RF) is particularly useful in that it allows a system to learn system-wide preferences, user specific tastes or those related to particular activities. It can take very little effort from the user to provide useful data to the system (Heesch and Rüger ; Rui, Huang, and Mehrotra 1997; Meilhac and Nastar 1999; Celentano and Chiereghin 1999; da Silva, Barcelos, and Batista 2006) making the time spent on providing it less onerous to the user.

The excellent overview of the use of relevance feedback systems of Ruthven and Lalmas (Ruthven and Lalmas 2003) is particularly useful and extensive. It describes the evolution of simple boolean models to more complex vectors space, probabilistic and logical models along with their individual strengths and weaknesses. Its conclusions demonstrate the stability and general utility of RF and while it was focused on text based search, many of the techniques can be directly transferred to other media types. It also wisely highlights the point that RF on its own is not sufficient to improve the search process and that it is only a very useful component to be used in conjunction with others - it is no panacea for an existing, ineffective system.

The work of Teevan *et al.* (Teevan, Dumais, and Horvitz 2005) widens the perspective on feedback to look at the past interactions of a user with a search system to develop a personalised model of their behaviour that can be used to improve future interactions by influencing the ranks of returned results. In contract to Heesch *et al.* (Heesch and Rüger), Teevan *et al.* showed that explicit feedback was not required to significantly improve the performance of the test systems in question. While their work was conducted solely on text based data sets their ideas can also apply to other types of media. While predominantly comprehensive, Teevan does use external search systems (in this case Google) as part of her search process. As this is a closed process whose working is not fully understood by the community, any improvements shown by her work cannot be said to be categorically down to their methods alone, but possible down the the unknown techniques of the external engine. Their framework of behavioural model building is nonetheless useful and transferable.

Feedback is very closely related to the physical interaction a user has with a system and the work by Stejić (Stejic, Takama, and Hirota 2003) demonstrates how the feedback process can be visualised and yet not alienate the user. It is also one of the first pieces of work to incorporate genetic algorithms into to the feedback process, with promising if inconsistent results.

2.7 Conclusions

In previous sections the key authors and their contributions pertinent to my work have been explored. One area lacking within the literature is that which fills the gap between the two established areas of social networks and their analysis and exploitation, and that of multimedia information retrieval. Whilst both these areas have been shown to have had extensive work carried out within them, there has yet to be much work in seeing how the one can influence and constructively add to the other.

The data gathered by Negoescu (Negoescu) and by Ortega and Aguillo (Ortega and Aguillo 2008) will be an invaluable starting point for further investigations in to the Flickr social graph. Whereas their work is limited in the scope of inter vertex connections, it nonetheless provides promising indicators to the general structure of the system which could later be exploited. Both of their results suggest the social network made up of 'Contact' relationships associated with Flickr is small-world and Negoescu's suggest it is scale-free, which is particularly promising.

Although these previous studies have only been carried out on social network made up of one type of relationship out of the many that exist, their results are still important as they show that at least some part of the system is both likely to be small-world and scale-free. Work still needs to be carried out on the other forms of relationship, and how the social networks they form interact. It is possible that the information gained by analysing combined social graphs yield more useful data than the individual graphs, just as multimodal search engines that combine evidences from different sub-systems can perform better than the individual sub-systems.

Implications of the existence of scale-free networks within Flickr This would imply that predictions could be made about future growth and the effects of node addition and removal can be modeled using existing techniques. It also means that small sampled datasets can be produced that can be made to represent the overall characteristics of the whole network of this type.

Implications of the existence of small-world networks within Flickr This implies that most users are in close social proximity to most other users with respect to the'Contact' relation. This would mean that there would be a lot of over-lapping in explicit user groups and quite possibly in implicit clusters as well. This does not exclude the possibility of deriving groups that discriminate users well.

In terms of methodologies highlighted by this review, one issue arises again and again. The gathering of data and the quality of the resultant data sets is particularly important and highlights the need of the community to establish some common sets to work with. There are currently no very large, representative data sets that incorporate multimedia with social data available to the community and this limits the evaluation that be undertaken by peers and makes judgements of competing methods difficult to make.

The connection online behaviour and social context is frequently assumed but not, I feel, adequately justified. Not enough work has been carried out on firmly establishing this connection and fully exploring it.

While the area of large scale online social data analysis has only been around for a few years due to the recent technologies that have enabled it, research is being carried out throughout the domain. My work should complement existing threads of research whilst building on and improving existing work through extension, validation and further exploratory analysis.

Chapter 3

Research question elucidation and motivation

3.1 The Research Question

After having analysed the related work connected to my area of research I have formulated the following question that sums up the direction of my investigation:

> *"How can the influence of local proximity social network data be modelled so as to enable a browsing/search experience more relevant to the end user of a large-scale online media sharing system?"*

As has been shown by my literature review, there has been little work on the interaction between social context and multimedia IR. Although personal behaviour online has been shown to be beneficial (Teevan, Dumais, and Horvitz 2005), the full extent of the usefulness of social data has yet to be explored. My work will continue to investigate this area and show whether and how social context information is useful within this field. In particular my work will be focussed on image retrieval in systems like Flickr.

Relevance as described in my research question shows the metric that I shall be investigating. As opposed to pure precision/recall measures commonly used in text based Ir and IR in general, due to the social nature of the interactions users have with online photo sharing web sites I believe there are other, more subjective criteria for judging how good a result set is. This also introduces the idea that the quality of results sets is subjective and cannot be entirely described by machine measurable statistics.

3.1.1 Hypothesis

My hypothesis could be described thusly. Giving a search system more data has been shown in my related work section to allow it to make more informed decisions regarding relevance when searching a data set. Social data is not extensively used by large, online media sharing sites. By looking at how social attributes and inter-relationships can be modeled and then incorporated into the retrieval process, the relevance of search results can be increased.

Although there exist attempts to formalise image content metadata (MPEG-7 is one example) there is currently no accepted model that describes the social context of a user or their interactions with online systems in a way that allows this data to be exploited or to be shared with other systems that could make use of the same data. This lack of a common language to describe users in an online context hinders the development of tools which can take advantage of this useful type of information.

In order for a model to be formulated that describes a user in terms of their attributes and connections to other users, the types of attributes must be analysed and work must be carried out to better understand the types of connections between users. As will be shown in Chapter 4, users interpret online systems in different ways and any model built to encompass such relations must take this act of interpretation into account.

This user model could be extended to describe the nature of groups of users and the attributes they exhibit as a community that are not evident on an individual basis. By doing so, group behaviours can be analysed and exploited to provide functionality that would not be possible by using only the context of users in isolation. The identification of both explicit and implicit groups of users

It must be emphasised that online social networks are inherently dynamic and are consistently changing to reflect the needs of those who contribute to the data of which they comprise. Any model built to encompass social context data will need to be capable of formalising this changing nature of user information. Some way of recording past behaviour would be useful in providing context for current activity which could help systems to more accurately interpret the information requirements of its users. By being able to model the evolution of a user's attributes and behaviour, systems would be able to avoid making inappropriate search algorithm decisions based solely on snapshots of user context.

The alternative to my hypothesis would be that the implication that user data can improve search systems doesn't hold and that the use of social data will either have no effect or degrade the relevance of search results. My contingency planning in Chapter 5 take these possibilities into account.

I intend to hone this hypothesis during my experimental period of my second year depending on the outcomes of my empirical work. If I discover a particularly fruitful area of results that would be worthwhile focusing on, my research question may change to reflect this.

3.2 Stake holders

The stake holders of my research and its impact are described in Table 3.1. By answering my research question I aim to satisfy the expectations of the different groups shown in that Table. These stake holders will be involved in different way during my research, from personal involvement to the different teams I shall be working with. It will be important to keep in mind what these groups are expecting from my work.

	Literature Review	Experimental Work	Conclusions
Myself	Familiarisation with the field is vital and the process of producing the review will help consolidate my knowledge	The development of thoroughly, robust, justifiable experiments will be an important part of my research skills improvement	Personal satisfaction at being able to produce justifiable conclusions
Academic Community	If extensive and well reasoned will be particularly useful to others who require an overview of the field	The results I gather regarding Flickr structure and user image trends will help support existing data available in the community, especially any datasets or novel methods I develop	My work could become a starting point for further research in this area
Industry	As with the academic community, a good review will provide industrial researchers a useful map to the field	Experimental results may give quantative metrics for companies to test their systems by	My results may have commercially exploitable outcomes in terms of improving user experience or system performance
General Public	Ultimately benefit by having better search systems based on my research	Online community web service developers may find my results useful	If my work results in better performing or more enjoyable search systems, the public will be able to directly take advantage of this

Table 3.1: Stake holders in my research

3.3 Completion Metrics

In order to know whether I have satisfied my research question I will test my work against the following three points:

Do I know more about the nature of social networks that coexist with multimedia datasets than the community did before I started my research? As shown in the literature review of this report, some initial studies into specifically this area have been carried out, but to few to really draw justifiable conclusions. If my work corroborates (or at least extends) that of existing work in a manner that is acceptable to the community, this criterion will be satisfied.

Have I been able to demonstrate how social networks can improve multimedia information retrieval? The analysis of my planned experiments I hope will lead to discovering exploitable trends that can be shown to be effective in improving IR. I will need to show either quantative and/or qualitative improvement through automated and user testing over existing systems.

Has my work been exposed to and tested by the academic community to the point where it is accepted? By comparing my work to the community of researchers working in the same field I shall be able to more easily judge its significance. In order to do this I will have to interact with the community continually to good effect.

Chapter 4

Experiment Pilot Studies

In order to establish the possible scope of future experimental work, initial pilot experiments were devised and the resultant methods and data are analysed in this chapter. The experiments undertaken here were to help give some indication as to issues that will need to be addressed before the larger, more comprehensive experiments will be able to be carried out when easier access to the data will be possible. Overall, the experiments were designed to give indications as to whether there are exploitable trends in the image data associated with the users that make up the Flickr social network. The experiments were designed with the quality of their methodology as an utmost priority.

4.1 Objectives

The following are the objectives for my experimental pilot studies. Some are aimed at building up knowledge about the network in question and their associated media, others at developing software and systems that will also be useful in future experiments.

1. To gain a better understanding of the systems used to access data from the Flickr system and to find out which algorithms are most appropriate to handling its data.
   Flickr's entire database of publicly accessible data is available via its API (Application Programming Interface). This allows programmes to be developed that can access data quickly and directly.

2. To develop experimental frameworks that will scale up as required for future work that function effectively and efficiently.
   Methods of data acquisition have to be tailored specifically to the need of the subsequent data analysis. Once these needs are established, a framework made up of robust and flexible code, algorithms and processes can be developed that can be used again in further experiments, establishing repeatable experimental methods that can be scrutinised by the community.

3. To gain a better understanding of local scale characteristics of users within the system.

There is currently no published data as to average connectedness within the different layers of the Flickr social network. Other values like the average number of explicitly judged favourite images is also not known external to Yahoo!

4. To find out, on average, how similar explicitly relevant images are for each user
If a user's favourite images are similar it would imply that a user's tastes or behaviour can be modeled to allow a system to provide results consistent with these trends. Users may interpret the label 'Favourite' differently and therefore use the label differently, giving sets of images that vary in different ways with respect to their similarity.

5. To find out how discriminating explicitly relevant images are between users
Although features for image sets may be extracted and compared, their ability to differentiate between users may to be insufficient for a system to exploit. If the user preferences discovered are too similar, it may be difficult to provide improvements.

6. To measure the change in similarity of images with respect to social distance[1]
If it can be shown that image set similarity reliable varies with respect to social distance, this information would give further credence to the hypothesis that those close to a user in their social network influence (or are described by) them.

4.2 Methodology

The graph of n users connected by their 'Contact' relations can be described as a graph where $G = (V, E, I)$ where every user is represented as the disjoint finite set of vertices $v_i \in V$ and each 'Contact' connection between two users is described as $e_{i,j} \in E$. I is the incidence relation that means that every element in E is incident to exactly 2 distinct elements of N and no two elements of E are incident to the same pair of elements of V.

As 'Contact' is not a bidirectional relation between two users, $e_{i,j} \neq e_{j,i}$.

As the entire Flickr social network cannot be gathered for local analysis, subsets were obtained based on seed users who become the roots of tree graphs of their contacts. These seed users were gathered in a pseudo-random manner by selecting 10 users who had recently uploaded images to the site and repeating this 6 times at different times of the day at different points during a week to try and lessen the effect of geographic bias (users in different countries around the world are likely to be actively using Flickr at different times of the day). For each of the users, 40 of their favourite images were grabbed. This quantity was decided on as being able to give enough images to produce meaningful results without taking too long to gather. For each seed user, their contact network was crawled to a depth of 3 levels and each of the contacts of each user encountered had their favourites downloaded too.

[1] SOCIAL DISTANCE describes the minimum number of users relationships between a user and the target user. If A knows B who knows C and there is no direct connection between A and C, then the social distance between A and C is 2.

The algorithm implemented to gather the data was depth first, meaning it explored to the extent of the section of the social graph and began by downloading the appropriate data from there and working its way back up, for each user. Breadth first would also have been valid but as there was the possibility of not being able to finish a crawl to a specified depth due to encountering an extensive number of hub nodes, in a time constrained experiment I felt it was more important to get a vertical slice of the network in order to draw conclusions for the questions in the previous section regarding social distance. These could only be answered by having data at some distance from the root nodes.

This resulted in a set of 60 annotated tree graphs containing user information and favourite images where available.

Each tree was then divided into subsets for testing. Each subset described a tier in the graph, with a maximum number of tiers of 4 (the depth the Flickr Spider program crawled to). A tier describes all those users who are an exact distance form the root user. A node can be said to be in tier T^x if it satisfies:

$v_{i,j} \in T^x \leftrightarrow \min\{n|A^n[i,j] \neq 0\} = x$,

where x is the distance away from the root node, A is the adjacency matrix of size $n \times n$ of graph G. The tiers are all subsets of the full graph such that

$G = \bigcup\limits_{k=0}^{t} T^k$, where T^0 describes the root node.

Each user's tree graph was then split into sub-trees made up of combinations of the different tiers; 1, 1+2, 1+2+3 in order test whether intra-group similarity decreases with the inclusions of users who are further away from a user with respect to their social graph. The images associated with each sub-tree were then analysed.

Six features were extracted from each image. The features were chosen to cover the three main feature types; colour, shape and texture. Their parameters were picked based on recent experimental work of other members of my research team.They were the MPEG-7 HDS colour descriptor, HSV colour descriptor, image thumbnail, and the smoothness, uniformity and Tamura texture descriptors.

The resultant feature vector of descriptor information could then be compared. The Manhattan (also known as the city block) distance was used as it has been shown to be high performing in similar circumstances.

Distances between feature vectors were then calculated between all the favourite images of a user (where more than one were available) giving an average dissimilarity value. This value characterises this group of favourites and could be used to compare to other users favourite groups.

The average of these dissimilarity values was then taken over the sub-tree to give an indication of how similar images were for different sections of the social network. It was hoped that as more data was included from further away in the network, the similarity would decrease, following the idea discussed in the Related Work Chapter regarding the effect a user's social network has on their behaviour.

Control data sets, In order to find out if the groups of favourite images explicitly selected by a user have any visual features in common, a control set of random images was required to compare against.

	Seed users	All users	Negoescu	Ortega and Aguillo
Number of users	55	26,645	3,544	663
Average number of FAVOURITES	60.03	23.22	-	-
Average number of CONTACTS	29.16	261.34	123.85	279.50 §
Average percentage of PRO users	73%	-	-	54.25% *
Average inter-nodal distance	-	-	3.88	2.8

Table 4.1: Data set statistics. Not all values were calculated for each data set. The value marked * was calculated from values in the work of Ortega and Aguillo. The value marked § is for the entire data set, not just for the 'large cluster' that was the focus of their work.

	1	1 - Control	1+2	1+2 - Control	1+2+3	1+2+3+4
HDS	29.42	9.50	24.25	10.79	22.47	24.07
HSV	13.85	3.77	10.58	4.62	12.65	12.13
Thumbnail	15.80	7.14	12.25	6.10	14.53	13.79
Smoothness	26.23	14.33	22.32	16.01	23.68	23.69
Uniformity	84.88	98.16	74.00	70.57	85.71	76.50
Tamura	24.47	12.96	21.15	15.34	22.94	22.31

Table 4.2: RSD values for the different data sub-sets, both gathered data and the control sets.

Representative datasets were produced that had the same number of users and the average number of images per user as a tier set. Images were selected from Flickr in the same pseudo-random manner as they were the for tier sets.

4.3 Analysis

In total, over 72Gb of images were downloaded to form the data set comprising of details of 26,645 users and 618,769 favourite images of these users sourced from the initial 60 pseudo-randomly selected users.

The resultant sets of group dissimilarity values were then statistically compared. As each feature that was extracted has different distribution and scales for values, the data could not be directly compared. As I could not guarantee that the distribution for each feature I used was Gaussian[2], for the moment, I decided to only adopt methods that dealt with non-parametric data.

The Relative Standard Deviation (RSD) was calculated for each feature for each sub set, see Table 4.2, which is the coefficient of variance represented as a percentage:

$RSD = \frac{\sigma}{\mu} \times 100$

The RSD gives an indication as to how discriminatory a feature is. A higher ratio would imply a feature has a more dispersed distribution about its mean and a smaller value would imply that the feature does not vary much for the given set of images.

Figure 4.1 shows the distribution of the dissimilarity values for each feature

[2] The data could be subjected to the Normality test, such as Kolmogorov-Smirnoff, to judge how normal my distributions are. This is planned for near-future work.

and for each tier as social distance increases. In Figure 4.1(d) it can be observed that the 'Smoothness' difference distribution shifts to the right as the social distance increases (as further tiers are included in the dataset), implying that as social distance increases, the difference between images also increases. The other distributions show little if any major change as social distance increases, implying that these features are not sensitive to the changes inherent to increasing social distance in the datasets, or that the reason for the 'Smoothness' shift is not down to social distance.

Figure 4.2 shows the frequency distributions for the six feature tested compared to control sets of random images. For all the distributions, it can be seen that the mean tends to stay the same but the variance changes between the data and control set, as exemplified in Figure 4.2(a). For each feature, the variance is less for the control set. This would seem to imply that a user's 'Favourite' images are more diverse in terms of the visual characteristics than random images with respect to certain image features.

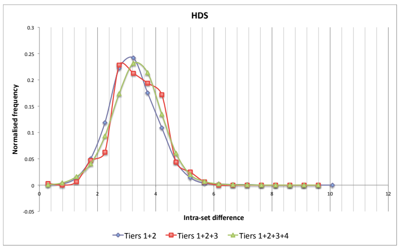
4.4 Evaluation

The pilot study experiments are best evaluated with respect to the initial objectives:

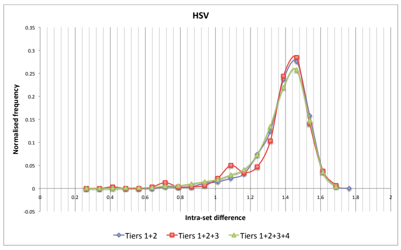
1. To gain a better understanding of the systems used to access data from the Flickr system and to find out which algorithms are most appropriate to handling its data.
   During the course of building software to interact with Flickr, the data structures used by the system and their methods of access were fully explored. Any future use of Flickr will be far more effective.

2. To develop experimental frameworks that will scale up as required for future work that function effectively and efficiently.
   The system for crawling through the Flickr network was built and tested in trials of obtaining a few hundred users to a correct social distance by using the Flickr API. This is a robust framework that can be further built upon in future work. The feature extraction software, partly built using existing tools performed well while handling the occasionally inconsistent data from Flickr. While it was not able to process all the data gathered as quickly as hoped, this was more down to optimistic expectations on my part regarding processing resources available.
   Any future experiments of a similar nature will be based on the processes built up for this pilot study.

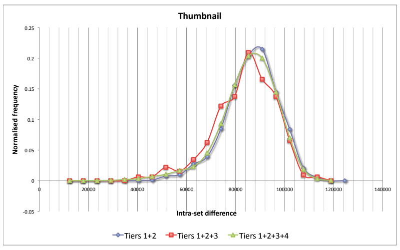
3. To gain a better understanding of local scale characteristics of users within the system.
   The values regarding the number of favourites, contact, images uploaded and whether users had paid to subscribe to the system were calculated for most of the subsets of the data gathered. As some of these values required a number of API calls to make, they were not calculated for some of the larger sets, as this would have had a major impact on the crawl time. The values gathered are compared to the number calculated by researcher in previous work in Table 4.1. It can be seen that for the



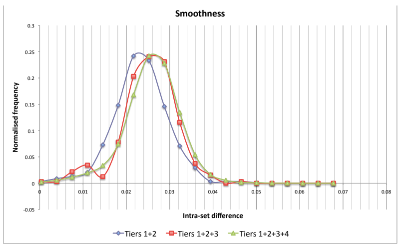
(a) HDS



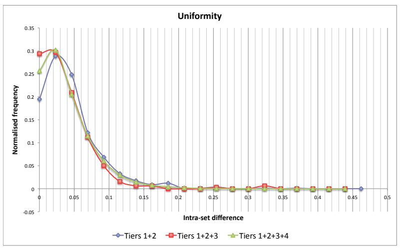
(b) HSV



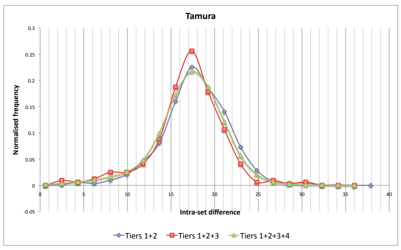
(c) Thumbnail



(d) Smoothness

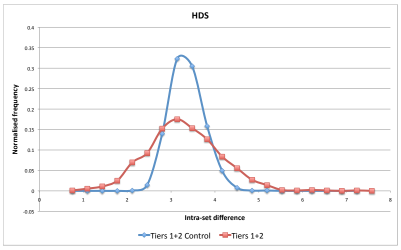


(e) Uniformity

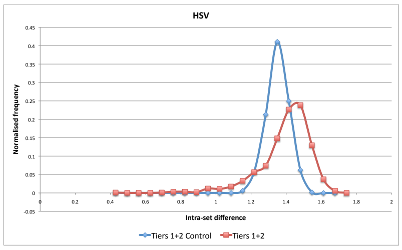


(f) Tamura

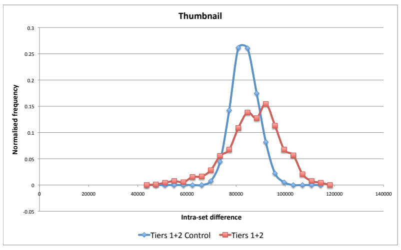
Figure 4.1: The effect of social distance on intra-set image dissimilarity as demonstrated by six image features.



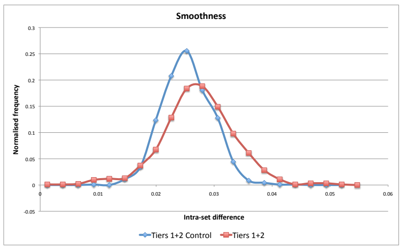
(a) HDS



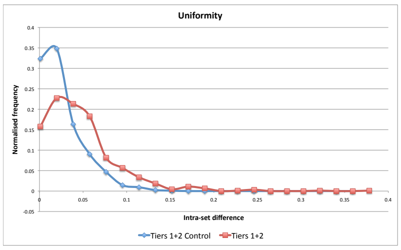
(b) HSV



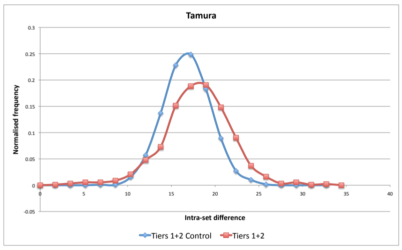
(c) Thumbnail



(d) Smoothness



(e) Uniformity



(f) Tamura

Figure 4.2: The diversity of images within a user's 'Favourite' set compared to sets of random images. It can be observed that the variance of the control set is smaller than that of the gathered data.

larger of the pilot study datasets, the connectivity of users in terms of their Contact relationships is of the same order and similar to that of the work of Negoescu (Negoescu) and Ortega and Aguillo (Ortega and Aguillo 2008). This helps support the representativeness of my data set in terms of its graph structure which implies that conclusions drawn base on structure are in some way justifiable.

4. To find out, on average, how similar explicitly relevant images are for each user
   The sets of images associated with different users were compared and the results are shown in Figure 4.2. It was observed that the images a user explicitly annotates as relevant end to be more dissimilar than sets of random images. This is particularly important, as it shows that in terms of image dissimilarity there is a trend, but whereas one might intuitively think that a user's 'Favourites' would be more similar to each other than a set of random images, this does not, according to this pilot study appear to be the case.

5. To find out how discriminating explicitly relevant images are between users
   Figure 4.2 also shows that some features help differentiate between data sets more easily than others. While the variance is higher than random for the HDS feature in Figure 4.2(a), Figure 4.2(f) shows that the Tamura feature is less discriminating. It also appears that the colour based features tend to be more discriminating than the texture based features, making their class of feature more suitable for characterising a user's favourite images.

6. To measure the change in similarity of images with respect to social distance
   The data shown in Figure 4.1 demonstrate that with increased social distance, the mean average difference between images in a social network tier increases. Although larger tiers have not yet been included here as their processing has taken longer than expected, primary analysis of this new data indicate a similar trend. This implies that users favourite images are influenced by their social network, which helps justify the further investigation into how and why, as well as the production of tools and systems that exploit this information.

4.4.1 Implementation Issues

Algorithmic decisions The initial decision to use a depth first search was perhaps not the most appropriate choice as it meant that when time was restricted and a crawl had to be halted, the resulting dataset would have incomplete tiers instead of missing tiers. If time and resources were not an issue, this wouldn't matter as either form of algorithm would be sufficient. As it is, data sets gathered using the methods outlined in the pilot study could be left biased if finished early.

4.4.2 Data Issues

Question of randomness The pseudo-random method used to gather images randomly where required by the experiments was imperfect as due to the method employed, it was biased towards users with recent online activity. As access to data was limited to that available through the API, a more random methods would have been difficult but not impossible. A better approach would have involved making more time-expensive calls to the Flickr servers, and it was ultimately decided to the method chosen as a compromise.

Bias Mitigation Geographic bias was lessened by selecting users at different times of the day on different days of the week to endure that the selection process did not favour users awake and active online at certain times (and therefore certain places). While effective, some form of geographical bias cannot be ruled out.

Representativeness The data set gathered was flawed in one way with respect to structural representativeness in that only users who were connected to each other via a 'Contact' relationship were gathered after selecting the initial root users. The resultant dataset, while connected, cannot be said to represent the structure of Flickr as a whole, as singleton users and disjoint user clusters would not have been included. As the focus for the study was for the images associated with the social network that was crawled, this is not a major concern, but does affect conclusions made about local graph structure.

Also, as not all users define favourite images, and probability distribution associated with having or not having favourites with respect to social distance is unknown, the analysis regarding image similarity and social distance is also likely to be biased in some small way.

4.4.3 Analysis Issues

The analysis so far carried out on the data set gathered is relatively crude (the use of the Relative Standard Deviation for example) but serves to indicate trends that were otherwise unknown. Future work may include a more detailed analysis that can measure correlation between data sets more comprehensively, particularly to measure Normalised Mutual Information.

Chapter 5

Further Planning

5.1 Time Planning

The first year of my PhD is described in Figure 5.1 and shows that most of the work planned for the year has been completed except for some remaining tasks that make up my research group's participation in evaluation workshop ImageCLEF 2008. Provisional plans for work to be carried out in the next two years are given in Figures 5.2 and 5.3. These are designed with flexibility in mind and I have 6 months of time available to use as a contingency if I find out certain tasks take longer than expected.

In order to expand upon the work of my pilot studies I will initially familiarise myself with the data and other resources available to me at the Yahoo! Research labs in Barcelona. I will then be in a position to more accurately plan the experiments I will need to undertake in order to investigate my theories and provide evidence to support my final conclusions. It is, however, imperative that I do as much of my experimental work as possible during my months in Barcelona as this will become more difficult to carry out remotely when I return to the UK. For this reason, the first months of my time there will be devoted to my planned experiments.

Although most tasks are dependent on their predecessors, if the need arises, some work can be carried out in a different order and the plans with be adjusted.

My primary deadline is 3.5 years after my start date, but I at present I am aiming to finish within 3 years to give myself some flexibility and scope for extension of promising experimentation and to allow for some contingency if my estimates for task duration are inaccurate.

The plans laid out in this Chapter will be revised every 3 months to establish progress, to ensure that targets will be met and to find out whether any adjustments need to be made.

5.2 Data Set Construction

More complete data sets for community analysis Although I will have access to Flickr snapshots while hosted at the Yahoo! Research Barcelona labs, there will be issues as to how much information can be made public and in what way. In order for my work to be accepted by the community (see criteria in Section 3.3) my experiments need to be repeatable and verifiable. For this to be possible, the data I work on needs to be available in some way.

Although I have developed a data set during my first year it is not of high enough quality for more comprehensive analysis. I shall therefore need to produce new datasets using the knowledge and tools I developed while producing those from my first year of work.

With access to snapshots of the entire Flickr system, the production of representative samples should be much easier. Also, it might be possible to run some experiments over the entire system if access methods allow and processing resources are available.

During the course of my pilot studies I came to understand the difficulties of scaling up image-analysis based experiments and the issues of computation time. Any model that I develop based on expressing the context of a user online, especially with respect to their interaction with media sharing systems will need to computationally feasible. The data sets that I will develop in order to help formulate my model must therefore reflect the nature of exisitng media sharing systems but be of an appropriate scale to allow feasible experimentation. The lessons learnt from my pilot studies will go towards limiting over ambitious attempts at too large a scale of data set generation, whilst maintaining representativity.

As my research question is focussed on the development of models to better describe user context online, the data sets I generate must reflect the dynamic nature of the systems under scrutiny by allowing some analysis of the evolution of user attributes, preferences and behaviour. This will entail either taking multiple snapshots of carefully selected sub-sets of users, or by continuously monitoring these users and maintaining an index of all changes.

Also, as my research question makes reference to user context within the context of large-scale media sharing systems on the Internet, I must ensure my data sets are as representative as possible. This will be done by using the information gathered both by myself in my pilot studies as well as other work in the field on the nature, structure and development of social media systems.

5.3 Planned Deliverables

I intend to develop the following resources based on reasoning from the work I have carried out so far:

A comprehensive analysis of the structure of Flickr I intend to gather enough data to give a detailed description of the various social networks that make up Flickr. This will be done by using the software and knowledge build up during my pilot studies. I also intend to find out how the different social relationships transfer into graphs and how they in turn affect user behaviour by further analysing trends in image similarity between different sets of users.

I will build models that help describe how a user interacts within the different type of networks made up by the different social relationships possible in the Flickr system. I will attempt to gain a fuller understanding of how the time dynamic nature of the social networks of Flickr affects the interaction behaviour models of its users. This will involve both analysing static data sets but also developing systems to monitor the change of Flickr over time.

A comprehensive analysis of the images of Flickr First of all I wish to validate the conclusions drawn from my pilot studies regarding the similarity of images explicitly defined as relevant. The counter-intuitive trend highlighted needs to be further investigated and its implications for systems that use this finding should be more fully explored.

I wish to develop current understanding of the images that are uploaded to Flickr in terms of their visual content and semantic context. This will involve continuing the work I started in my pilot studies but increasing the complexity of the data I gather. I wish to be able to categorise images with the system an in doing so build models that can characterise the data on Flickr as a whole data set.

I would like to introduce some semantic context data to the features I analyse as I think this will more accurately reflect the general context of the media in question and might give more discriminating ways of highlighting the way social networks affect user behaviour.

An interface to Flickr that can use social data in real-time to improve performance In order to test any theories that I develop regarding improvements to Flickr based on social context, I shall need to produce a system to test this both quantatively and qualatively. As Flickr is a dynamic system I foresee having to develop a system that can interact with Flickr and extract whatever data it needs in real- or near real-time. This would make the tool flexible and hopefully perform better by using the most up to date social data.
From the pilot studies I have so far carried out, it has become clear that there are system development paradigms that are particularly suited to experimentation, and others to production. While Java command line utilities were sufficient for gathering and analysing data, they would be less suitable for an implementation for end users. I have informally evaluated difference methods and I envisage a web based system based around AJAX[1] style sites that are both powerful in terms of handling data, flexible in that they can be used on many different types of computer system and provide a user friendly way of interacting with the system, without requiring too much client-side computation.

A demonstration system would allow me to easy and quickly show my work to others and make it easy to communicate my work, and so it is a priority for the next two years.

My thesis All the work I carry out over the course of my PhD will be described, justified and evaluated in my final thesis. My primary aim in producing this report is to demonstrate that I have developed my skills to become an independent researcher, having learnt and adopted appropriate knowledge and methods on the way. As it is such a vital deliverable, a significant amount of time will be devoted to it, as shown in Figure 5.3.

Publications In order to satisfy my criteria (Section 3.3) of having my work accepted by the academic community, I will have to expose my work externally to my research group. This will include involvement in conferences pertinent to my area, writing papers when possible and presenting my work to as wide a range of audiences as possible (from fellow academics to the general public).

[1]Asynchronous JavaScript And XML

5.4 Contingency Planning

My plans for experimentation and model building during my second year are dependent on the assumptions that the system I will continue to analyse will ultimately conform to the initial evaluation undertaken by myself and others as mentioned previously in this report. It is also defendant on the outcomes of work elsewhere in the information retrieval field that can be summarised as saying that personalisation improves the search experience with respect to certain metrics.

If the first assumption is proved to be incorrect and the Flickr system I will continue to investigate is somehow different to those systems that have been analysed more thoroughly in the past, this will not be hypothesis breaking problem. By understanding how Flickr works we will be able to refine our definition and understanding of such systems and should not limit my ability to model and improve on the system even if the structure or dynamics of the system are different to other systems. It would however require readjustment to my experimentation schedule as more effort would need to be invested in those studies that gather data on this area of understanding.

I believe it unlikely that the second issue of whether personalisation does ultimately improve IR should not more too troublesome as there is already extensive work (as discussed in my Related Work section) demonstrating its utility in particular contexts. I don't think the system and the user interactions I will be analysing will differ too greatly from this established areas of study.

If however I find that any model I build and system that exploits it does not yield any performance increase, I believe that the process of formalising and modeling online social context will still be very valuable to the community and my focus may change to reflect this aspect of my work.

5.5 Ethical Considerations

During my experiments to date I have only had access to data that has been submitted by Internet users in the knowledge that it would be publicly accessible. There are two issues I need to address that may occur during my experimental period.

The first is the issue of making use of data that a user has not explicitly sanctioned to be used by someone 'external' to the Flickr organisation. User logs are not publicly available but would be very useful in producing data sets that reflect the changing nature of online user interaction. It is also possible that they would not be necessary in order to derive th information I would require. Any and all use of user data would be in accordance with The Open University ethical guidelines in research.

The second is the slightly more difficult issue of the exploitation of information derived from publicly accessible data. If I am able to identify useful trends in a user's profile that they themselves may not be consciously aware of, am I justified in using this information if they have not explicitly given me permission to? Although i currently don't have a complete answer to this question I shall be keeping it in mind as my work develops. It may ultimately be a decision between limiting the data I use in the production of improved systems, or coming to a compromise justified by the ultimate benefits of any system I develop.

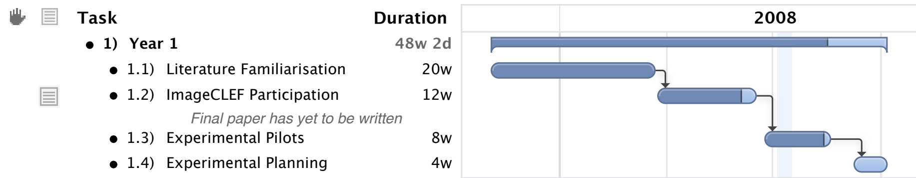


Figure 5.1: Year 1 plan of work. The darker bars indicated current progress for that task.

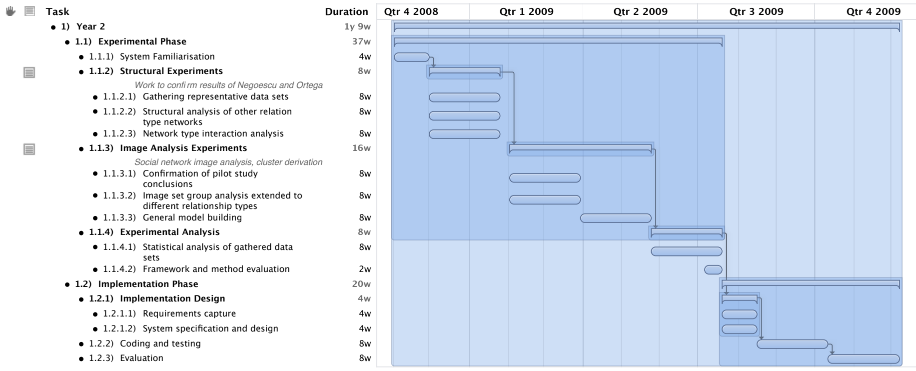


Figure 5.2: Year 2 provisional plan of work.

Any experiments undertaken with people would also be strictly done within the guidelines of both The Open University and Yahoo! Research.

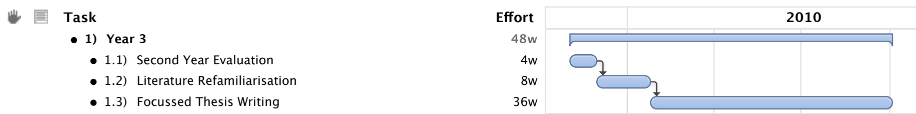


Figure 5.3: Year 3 provisional plan of work

Chapter 6

Achievements to Date

During my first 8 months of work I have been involved in a few other projects that have ultimately enriched my own work. Notable cases include:

Pilot Study. Part of my first year was dedicated to initial studies of the Flickr system as detailed earlier in this report. I have been able to draw some interesting conclusions from the data gathered as well as develop tools for gathering and processing data. Not only has it produced useful results, but has helped hone my experimental design, implementation and evaluation skills. It has been directly responsible for the direction of my planned experimentation of my second year and has helped in the clarification of my research question.

Data Set Production. I have also been able to produce a data set where there where none of its type was previously available to the community. It is also the first to include social network information along with image data. Its layout and extent is sufficient for initial exploratory work and it successors will be improved by using the lessons learnt during the first data set's production.

Participation in ImageCLEF 2008. The Multimedia Informations Systems (MMIS) Group of which I am a member entered 3 tasks of the annual ImageCLEF multimedia information retrieval evaluation workshop. These were focused on photo retrieval, retrieval of images from Wikipedia and a visual concept detection task.

My contribution was to build, test and use a system to combine multiple evidences from other members in the group to provide a final set of ranked images for submission. This involved getting familiar with the current treatment of multimodal systems and selecting the best and most appropriate existing methods for our own system. This was implemented and was shown, in certain circumstances, to provide better results than the individual evidences. Final performance results are pending, but a paper is being produced during at the moment for the ImageCLEF workshop at Århus in Denmark in November which I shall be attending as co-representative of our research team. Not only has this been a worthwhile research endeavour, but the project management experience acquired during the process has been invaluable.

Article in Inside Knowledge Magazine. An article entitled "A Picture is Worth a Thousand Words" was collaboratively written with Stefan Rüger and was published as a cover feature article in the May edition of the knowledge management magazine Inside Knowledge[1]. The article gave an overview of multimedia search methods and technologies and was aimed at those with moderate to little prior knowledge of the area.

Participation in PhD Conference 2008. The Centre for Research in Computing (CRC) at The Open University held its annual PhD Student conference in June 2008. I presented a position presentation as well as a poster. Not only was I able to expose my work and ideas so far to others in the field, but also form some very fruitful connections to academics with interests in my work and vice versa.

Involvement in group publications. I have helped in the submission of numerous papers from the MMIS group in my capacity as editor. This has been particularly rewarding as it has helped hone by language skills and improved my own writing and presentation.

Involvement in PhD Forum. The CRC runs a weekly forum for students of all stages to come together and develop their research skills and to improve networking between different departments. I have been a consistent attendee and have benefited immensely from both the skills based instruction and the social interaction.

[1] http://www.ikmagazine.com

References

Ahn, Y.-Y., S. Han, H. Kwak, S. Moon, and H. Jeong (2007). Analysis of topological characteristics of huge online social networking services. *Proceedings of the 16th International Conference on World WIde Web*, 835–844.

Aida, M., K. Ishibashi, C. Takano, H. Miwa, K. Muranaka, and A. Miura (2005, December). Cluster structures in topology of large-scale social networks revealed by traffic data. *Global Telecommunications Conference 1*, 6.

Arnia, F., I. Iizuka, M. Fujiyoshi, and H. Kiya (2007, September). Dct sign-based similarity measure for jpeg image retrieval. *IEICE Trans. Fundamentals E90-A*, 1976–1985.

Bartell, B., G. Cottrell, and R. Belew (1994). Automatic combination of multiple ranked retrieval systems. *proceedings of the 17th Annual International ACM SIGIR conference on Research and Development in Information Retrieval*, 173–181.

Baudrier, É., G. Millon, F. Nicolier, and S. Ruan. A new similarity measure using hausdorff distance map. *IEEE International Conference on Image Processing (ICIP).*

Bay, H., T. Tuytelaars, and L. V. Gool (2006, May). Surf: Speeded up robust features. *Proceedings of the ninth European Conference on Computer Vision.*

Cano, P., E. Batlle, T. Kalker, and J. Haitsma (2005). A review of audio fingerprinting. *Journal of VLSI Signal processing 41*, 271–284.

Celentano, A. and S. Chiereghin (1999). *Multiple strategies for relevance feedback in image retrieval.*

Chai, D., S. L. Phung, and A. Bouzerdoum (2003). A bayesian skin/non-skin colour classifier using non-parametric density estimation. *Proceedings of the 2003 International Symposium on Circuits and Systems 2*, 464–467.

Conole, G. (2001, August). Mapping information needs. *Advanced Learning Technologies*, 345–348.

da Silva, S. F., C. A. Z. Barcelos, and M. A. Batista (2006). An image retrieval system adaptable to user's interest by the use of relevance feedback via genetic algorithm. *Proceedings of the 12th Brazilian symposium on multimedia and the web*, 45–52.

de Jong, F. M. G., T. Westerveld, and A. P. de Vries (2007, March). Multimedia search wihtout visual analysis: The value of linguistic and contextual information. *IEEE Transactions on circuits and Systems for video technology 17.*

der Weken, D. V., V. D. Witte, M. Nachtegael, S. Schulte, and E. Kerre (2005). Colour image comparison using vector operators. *Proceedings of the 2005 Intenational Conference on Computational Intelligence for modelling, Control and Automation.*

Deselaers, T., D. Keysers, and H. Ney (2004, September). Features for image retrieval - a quantitative comparison. LNCS, Tübingen, Germany.

Dodds, P. S., R. Muhamad, and D. Watts (2003). An experimental study of search in global social networks. *Science 301.*

Enser, P. G. and C. J. Sandom. *Towards a Comprehensive Survey of the Semantic Gap in Visual Image Retrieval*, Volume 2728/2003 of *Lecture Notes in Computer Science*, pp. 163–168. Springer Berlin / Heidelberg.

Feng, Z.-H., B. Zhou, J.-Y. Shen, and D. Pan. Image comparison using symmetrical color-spatial feature. *Proceesings of the Fourth International Conference on Machine Learning and Cybernetics.*

Granovetter, M. (2008, March). Ignorance, knowledge, and outcomes in a small world. *Science 301*, 773–774.

Haitsma, J. and T. Kalker (2003). A highly robust audio fingerprinting system with an efficient search strategy. *journal of New Music Research 32*, 211–221.

Heesch, D., M. J. Pickering, S. Rüger, and A. Yavlinsky. Video retrieval using search and browsing with key frames.

Heesch, D. and S. Rüger. Performance boosting with three mouse clicks - relevance feedback for cbir.

Howarth, P. and S. Rüger. Evaluation of texture features for content-based image retrieval. *Proceedings of the International Conference on Image and Video Retrieval*, 9.

Hu, R., S. Rüger, D. Song, H. Liu, and Z. Huang. Dissimilarity measures for content-based image retrieval.

Ionescu, M. and A. Ralescu. Fuzzy hamming distance in a content-based image retrieval system. *Proceedings of the IEEE International Conference on Fuzzy Systems*, 1721–1726.

Jia, L. and L. Kitchen (2000). Object-based image similarity computation using inducive learning of contour-segment relations. *IEEE Transactions on Image Processing 9*, 80–87.

Kaijser, T. (1998). Computing the kantorovich distance for images. *Jounal of Mathematical Imaging and Vision 9*, 173–191.

Li, Z.-W., X. Xie, H. Liu, X. Tang, M. Li, and W.-Y. Ma (2004, October). Intuitive and effective interface for www image search engines. *MIR '04.*

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision.*

Manjunath, B. S. (2001). Color and texture descriptors. *IEEE Transactions on circuits and systems for video technology 11*.

Meilhac, C. and C. Nastar (1999). Relevance feedback and category search in image databases. pp. 512–517.

Mikolajczyk, B. and C. Schmid (2005). A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence 27*, 1615–1630.

Negoescu, R. An analysis of the social network of flickr. *Internal Publication - École Polytechnique Fédérale de Lausanne*.

News, B. (2008). Bbc news online - aboriginal archive offers new drm - 30/01/08.

Ortega, J. L. and I. F. Aguillo (2008). Structural analysis of an online social network: The spanish netowrk of flickr. *Profesional de la Información 17*.

Rui, Y., T. Huang, and S. Mehrotra (1997). Content-based image retrieval with relevance feedback in mars. pp. 815–818.

Ruthven, I. and M. Lalmas (2003). A survey on the use of relevance feedback for information access systems. *Knowledge Engineering Review 18*, 95–145.

Singh, S. K., D. Chauhan, M. Vatsa, and R. Singh (2003). A robust skin colour based face detection algorithm. *Tamkang Journal of Sceince and Engineering 6*, 227–234.

Stejic, Z., Y. Takama, and K. Hirota (2003, October). Relevance feedback-based image retrieval interface incorporating region and feature saliency patterns as visalizable image similarity criteria. *IEEE Transactions on Industrial Electronics 50*, 839–852.

Stevens, M. R. and J. R. Beveridge (2000). Image comparison techniques in the context of scene refinement. *Proceedings of the 15th International Conference on Pattern Recognition 1*, 1685.

Teevan, J., S. T. Dumais, and E. Horvitz (2005). Personalizing search via automated analysis of interests and activities. *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 449–456.

Vezhnevets, V., V. Sazonov, and A. Andreeva (2003). A survey on pixel-based skin colour detection techniques. *Proceedings of Graphicon 2003*.

Watts, D., P. S. Dodds, and M. Newman. Identity and search in social networks. *Science 296*, 1302–1305.

Watts, D. and S. Strogatz (1998, June). Collective dynamics of 'small-world' networks. *Nature 393*.