Evaluating the Implicit Feedback Models for Adaptive Video Retrieval

Frank Hopfgartner
Department of Computing Science
University of Glasgow
Glasgow, United Kingdom
hopfgarf@dcs.gla.ac.uk

Joemon Jose
Department of Computing Science
University of Glasgow
Glasgow, United Kingdom
jj@dcs.gla.ac.uk

ABSTRACT

Interactive video retrieval systems are becoming popular. On the one hand, these systems try to reduce the effect of the semantic gap, an issue currently being addressed by the multimedia retrieval community. On the other hand, such systems enhance the quality of information seeking for the user by supporting query formulation and reformulation. Interactive systems are very popular in the textual retrieval domain. However, they are relatively unexplored in the case of multimedia retrieval. The main problem in the development of interactive retrieval systems is the evaluation cost. The traditional evaluation methodology, as used in the information retrieval domain, is not applicable. An alternative is to use a user-centred evaluation methodology. However, such schemes are expensive in terms of effort, cost and are not scalable. This problem gets exacerbated by the use of implicit indicators. The use of a simulated evaluation methodology for the comparison of various interactive retrieval strategies has to be explored. In this paper, we explore the effectiveness of a number of interfaces and feedback mechanisms and compare their relative performance using a simulated evaluation methodology. The results show the relative better performance of an interface with the combination of explicit and implicit features.

Categories and Subject Descriptors

H.3.3 [**Information Systems**]: Information Search and Retrieval—*Relevance Feedback, Query Formulation, Selection Process*; H.3.7 [**Information Systems**]: Digital Libraries—*User issues*; H.3.4 [**Information Systems**]: Systems and Software—*Performance evaluation (efficiency and effectiveness)*; H.5.2 [**Information Systems**]: User Interfaces—*Interaction styles*; I.6.6 [**Computing Methodologies**]: Simulation Output Analysis

General Terms

Design, Experimentation, Human Factors, Measurement

MIR'07, September 28–29, 2007, Augsburg, Bavaria, Germany.

Keywords

adaptive video retrieval, implicit relevance feedback, user simulation

1. INTRODUCTION

With the improving capabilities of current hardware systems, there are ever growing possibilities to store and manipulate videos in a digital format, leading to the development of a number of video archives. People build their own digital libraries from materials created through digital cameras and camcorders, and use systems such as YouTube[1] and Google Video[2] to place this material on the web. Unfortunately, this data creation prowess is not matched by any comparable tools to organise and retrieve video information.

There is a need to create new retrieval engines to assist the user in searching and finding video scenes they would like to see from many different video files. Unlike text retrieval systems, retrieval on digital video libraries is facing a serious problem: The Semantic Gap. This is the difference between the low-level data representation of videos and the higher level concepts a user associates with video.

The semantic gap problem can be addressed to a great extend by applying techniques from the interactive textual retrieval domain. The visualisation of retrieval results and the design of interfaces are well studied fields in text retrieval. An important strategy in text retrieval is the query reformulation to improve retrieval results. A strategy to identify relevant results is the use of relevance feedback. Additional terms for query expansion can be gathered from the user's query and from relevant documents. The user's feedback and interaction with the system can be used to identify relevant results. There are different types of interactions, usually divided into two categories: explicit feedback and implicit interaction. Explicit feedback is given when a user informs a system what it has to do on purpose, such as selecting something and marking it as relevant. On the other hand, users' implicit actions can be interpreted as feedback. By mining implicit user interaction data, one can infer user intentions and thus could be able to retrieve more relevant information. An example is printing out a web page. This interaction may indicate an interest in the printed web page. The use of implicit feedback techniques in the textual retrieval domain has been studied extensively.

Bearing few instances within the TRECVID interactive tasks, the use of relevance feedback in interactive video re-

[1]http://www.youtube.com/
[2]http://video.google.com/

trieval has not been studied. One reason is that providing explicit feedback is a cognitively difficult process. Giving explicit feedback, users are forced to update their need, which can be problematic when their information need is vague [13] or when they are unfamiliar with the data collection [12]. Furthermore, users are lazy. They do not tend to provide much feedback on which to base an adaptive retrieval algorithm [6]. In addition, they are uncertain on how exactly such feedback will be used by the underlying retrieval system. It is also blamed for the lack of appropriate interfaces. Implicit feedback techniques, to an extent, address some of these problems. An implicit feedback approach tries to infer user intentions by mining user interaction data. It has been shown to be effective in the WWW [4] and the textual retrieval domain [15, 10]. However, implicit feedback techniques are not explored in the multimedia domain.

If used properly, these features are highly useful in reducing the semantic gap. Implicit indicators are often very noisy and hence, to make reasonable inferences about user intentions, we need to use a combination of them. This will lead to a number of possible combinations. In order to select and use the right combination of implicit indicators, we need to evaluate their performance. The only way to measure the effectiveness is through user evaluation, which is very expensive.

In this paper, we address these issues more specifically. We discuss the development of a simulated evaluation methodology for benchmarking interactive search interfaces and approaches. Our second focus is the study of implicit factors for the use in the multimedia search domain.

The paper is organised as follows: Section 2 gives a brief overview of existing video retrieval systems and discusses their inadequacies. In Section 3, we present the different feature combinations, which we divided into five user behaviour models. These behaviour models are introduced in Section 4. Section 5introduces our feedback weighting. In Section 6, we introduce the retrieval model our work is based on. Section 7 explains the need for a simulation framework. We introduce the simulation runs we performed in Section 8 and discuss the results in Section 9. Finally, we summarise the findings in Section 10 and discuss future work.

2. BACKGROUND

In this section, we discuss current approaches to interactive video retrieval. All of these systems are developed and evaluated within the context of TRECVID tasks.

Christel and Concescu (2005) [3] developed and compared two video retrieval systems using visual *and* textual data versus a visual-only system as part of the Informedia project. In addition, they compared expert and naïve users. Using their interface, users interacting with the combined system scored significantly higher on the performance metric of average precision. The expert runs outperformed the naïve users run. Their system includes a facility for explicit relevance feedback. Implicit relevance feedback, however, is ignored.

Foley et al. (2005) [5] developed a multi-user system using a DiamondTouch tabletop device. Using the interface, a user can add images as part of the query and select which feature of the image shall be a reference for similar results. They implemented two versions of that system: one with emphasis on efficient searching, the other one on increasing awareness of the users. They conclude that providing awareness cues improves the retrieval performance. However, their interfaces do not support relevance feedback. Search queries have to be refined manually without any automatic reference to former retrieved results.

Browne et al. (2003) [2] compared a video retrieval system based on text, image and relevance feedback with a text-only retrieval system. Precision/Recall of their runs show that the performance of both systems is comparable. However, comparing recall over time, the combined system outperformed the text-only system. Although they consider explicit relevance feedback, they ignore the information that can be gathered when considering implicit feedback.

Heesch et al. (2004) [7] experimented in video retrieval using searching and browsing with an emphasis on user interaction and user navigation. They developed two systems: one including both search and browsing, the other including search only. They conclude that adding the browsing functionality increases retrieval performance. The interface of their interactive video library retrieval system unites both visual and textual search queries and the ability of giving relevance feedback. Even though users can give explicit relevance feedback using their system, the knowledge which can be gained from implicit feedback is ignored [7, 9].

Hopfgartner et al. (2007) [8] introduce a model of implicit information for interpreting the user's actions with an interactive video retrieval interface. Based on a simulated user study, they conclude that their model seems to enhance retrieval results. However, the model is not advanced yet, the user behaviour assumptions are naïve and the approach rather primitive.

The above approaches are very similar: They use text and visual surrogates to identify relevant video shots[3], which are presented by keyframes. They are not comparable as the research results depend on different retrieval systems and interfaces. Hence, no indication on how to learn from their approaches can be extracted.

In text retrieval, both explicit and implicit relevance feedback techniques are seen as a appropriate approach to enhance retrieval results [15, 10]. However, it is also been shown that a combination of explicit and implicit relevance features may be useful to increase retrieval effectiveness [16].

To summarise, in video retrieval, the use of relevance feedback techniques are basic and exploratory in nature. Nobody has employed, yet, the concept of implicit relevance indicators in video retrieval. We, however, assume that a combination of explicit and implicit relevance features will improve retrieval results. Hence, first of all, it is our objective to study the *effectiveness of* the use of implicit features in video retrieval.

Our second focus is on developing a methodology on evaluating interactive video retrieval approaches. The traditional evaluation methodology, as used in the information retrieval domain, is not applicable. One reason is the lack of repeatability. Besides, it is hardly possible to benchmark systems using user-centred evaluations. Therefore, we introduce a new methodology in simulating user behaviour. The advantage of simulations are obvious: they are scalable, repeatable and, over all, cheap.

[3]A shot is a small fragment of a video. It is that part of the video, which is recorded using the same camera and the same angle. Shots can be detected automatically using visual features such as colour, shape and texture [1].

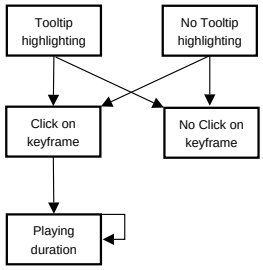
3. IMPLICIT INDICATORS

In this work, we will introduce a simulation framework which can be used to clarify whether implicit relevance indicators can influence retrieval results in a positive way. Therefore, a first necessary step is to emphasise valid implicit indicators. After analysing the log files and questionnaires of an interactive study conducted using an adaptive video retrieval system [14] and after analysing the interface approaches introduced in Section 2, we identified the following six implicit feedback categories:

- *Highlighting* when moving the mouse over a keyframe. This can result in a tooltip showing neighboured keyframes and additional text [8] or in highlighting the query terms in the text associated with the keyframe [2]. This feature indicates further interest in a keyframe as the user receives additional information about the result.
- *Click on a keyframe* to trigger playback of a video shot [2, 8] or to perform further actions [3, 7]. This feature indicates the users' interest in the video shot which is represented by the keyframe.
- *Using the sliding bar* to navigate through a video [2, 8, 3]. This feature indicates further interest in the video. Users appear to slide through a video when the initial shot is not exactly what they were searching for but when they believe that the rest of the video *might* contain other relevant shots. Hence, the initial shot might not be an exact match of the users' need but raises hope to find something of relevance in the same video.
- *Looking at metadata* (date of broadcast, broadcasting station,...) [8, 3]. Giving this implicit feedback, users show a higher interest in the current shot, as they want to get additional information. These information can help them to judge about the relevance of the shot. A user e.g. might search for a specific sports event such as the football world cup final. In such cases, the direct correlation between broadcasting date and event date can help to identify relevant shots as such events usually appear in the news shortly after their happening.
- *Browsing through a video* by clicking on its neighboured keyframes [2, 8, 7, 3]. Similar to sliding through a video, this feedback indicates users' interest in this shot. Unlike using the sliding bar, browsing indicates that users suspect a relevant shot in the neighbourhood of the current shot.
- The *playing duration* of a video indicates users' interest in the content of the video.

These interactions can be used for implicit relevance feedback and be employed in different combination. We do not define whether these interactions are positive or negative indicators for relevance. The interpretation and the importance of these indicators depend on the interface context and also the underlying retrieval model. However, it would be impossible to evaluate these combinations for effectiveness using user-centred evaluations. Therefore, one focus of our work is to establish a methodology for evaluating adaptive video retrieval systems.

Figure 1: Possible combinations of I_1



We modelled five different user behaviour scenarios, introduced in Section 4, which support different combinations of these implicit features. They model possible user behaviour using the interfaces presented in Section 2. Using these models, we ran a user simulation to clarify the influence implicit indicators can have on video retrieval. These results will shed light on the possibility of implicit factors for retrieval.

4. USER INTERACTION SCENARIOS

A useful way to identify the reliability of implicit features is in testing the effects of different combinations on its retrieval results. The different state-of-the-art interfaces introduced in Section 2 can use various implicit features that can be adapted for implicit relevance feedback. They are summarised in Section 3. To study the effect of implicit features in retrieval, we analysed possible interactions between users and some of these interactive video retrieval systems [2, 8, 7, 3]. Based on this, we modelled five possible user interface scenarios. To keep the experimentation simple and comparable, we do not integrate every feature provided by each interface. Hence, a scenario covers *possible* user interaction, not necessarily a user interaction including *all* features the interface provides. This guarantees the combination of different implicit features in our models. Various interfaces result in different user's interactions with the interface and thus, trigger different implicit relevance feedback. The following user interface models I_1 – I_5 afford different user interactions and search strategies.

4.1 User Scenario 1 – I_1

In this user interface scenario 1 (I_1), the system presents the results of a query where results are presented by keyframes as in [3, 8] and provides tooltips when the user moves the mouse over the interface. It also includes playing the video shot.

In our model, a user will (i) move the mouse over listed keyframes to get some additional information of the shot in a tooltip. Based on this information, the user may (ii) click on the keyframe to (iii) start playing a video.

These actions result in the use of the following implicit relevance feedback:

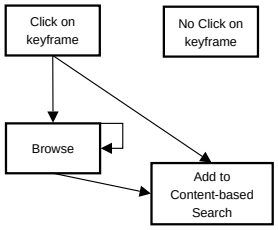
i. Highlighting of a tooltip

ii. Click on a keyframe to trigger video

iii. Playing duration

The possible behaviour combinations are visualised in Figure 1.

Figure 2: Possible combinations of I_2



4.2 User Scenario 2 – I_2

The user scenario scenario 2 (I_2) models possible feedback that can be given when using the system introduced by Heesch et al. [7]. In this interface, information will be presented on different panels. Retrieval results are presented by keyframes. Clicking on one keyframe in a result panel will set focus on that keyframe and update all other panels. One panel contains the neighboured keyframes in a fisheye presentation. In this panel, a user can browse through the results. In every panel, the user can right-click on a keyframe and add the frame as content-based (visual) query.
In this model, users (i) click on a keyframe in the result list and (ii) browse through its presented neighboured frames. They can always right-click on a keyframe and (iii) add the frame as visual query. These actions result in the use of the following implicit relevance feedback:

i. Click on a keyframe to update panels

ii. Browsing through neighboured keyframes

iii. Add keyframe as content-based query

Possible behaviour combinations are visualised in Figure 2.

4.3 User Scenario 3 – I_3

The third user interface scenario (I_3) covers a behaviour which can be achieved when using the text-only video retrieval system provided by Browne et al. [2]. Their web interface ranks retrieved results in a list of relevant video programmes. Each row displays the most relevant keyframe, surrounded by its two neighboured keyframes. Below the shots, the text associated with the result is presented. The query terms which are associated with the keyframe are highlighted when the user moves the mouse over the keyframe. When clicking on a keyframe, the represented video shot can be played.
In our model, a user can (i) click on a keyframe to trigger (ii) video playback. The user can (iii) highlight associated query terms and (iv) navigate through the video using a sliding bar.

These actions result in the use of the following implicit relevance feedback:

i. Click in a keyframe to trigger video playback

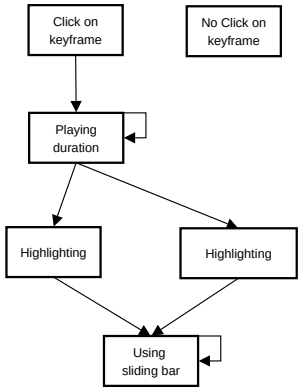
ii. Playing duration

iii. Highlighting associated terms

iv. Using the sliding bar

Possible behaviour combinations are visualised in Figure 3.

Figure 3: Possible combinations of I_3



4.4 User Scenario 4 – I_4

This model simulates the user's interaction with the system provided by Hopfgartner et al. [8]. In their interface, retrieved video shots, represented by a keyframe, are listed in a result panel. Moving the mouse over a keyframe will highlight a tooltip showing its neighboured keyframes and the associated text. When clicking on a keyframe, the corresponding video is played. Additional surrogates such as broadcasting station and date can be highlighted when moving the mouse over the video. The video which is currently played is surrounded by its neighboured keyframes. A user can click on them and browse through the current video. Also, a user can use a sliding bar to navigate through the video.
In this model, users can (i) highlight additional information in moving the mouse over a retrieved keyframe to get some additional information of the shot (neighboured keyframes and text from the speech recognition software), (ii) click on a keyframe of a result list and (iii) play a video. Besides, they can (iv) browse through the video to find new results in the same video.

These actions result in the use of the following implicit relevance feedback:

i. Highlighting of a tooltip

ii. Click on a keyframe to trigger video

iii. Playing duration

iv. Browsing in a video

Possible behaviour combinations are visualised in Figure 4.

4.5 User Scenario 5 – I_5

This user interface scenario I_5 is based on the retrieval interface by Christel and Concescu [3]. In contrast to the other scenarios, it supports explicit relevance feedback. In this interface, retrieved results are represented by keyframes and presented in a list. Clicking on one keyframe, the user can choose to explicitly mark a shot as relevant, to play the video, to show a storyboard or to display additional information. The storyboard will list keyframes in a chronological order.
In this model, users (i) click on a keyframe in the result list and (ii) play a video. They can also (iii) use the sliding bar.

Figure 4: Possible combinations of I_4

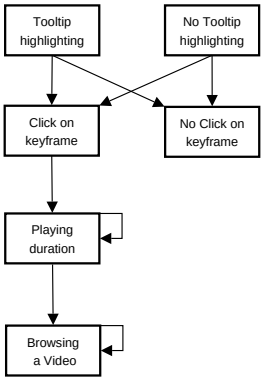
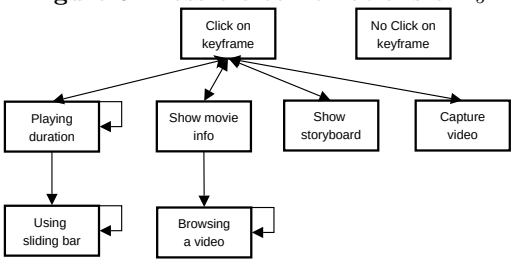


Figure 5: Possible combinations of I_5



Users may open the storyboard and (iv) browse through the video to find new results in the same video. Moreover, they can (v) show additional movie information and sort results by date and broadcasting station. Besides, they can explicitly judge the relevance of a video shot.

These actions result in the use of the following implicit relevance feedback:

i. Click on a keyframe to trigger video

ii. Playing duration

iii. Using the sliding bar

iv. Browsing in the storyboard

v. Listing of date and broadcasting station (metadata)

Possible behaviour combinations are visualised in Figure 5.

5. FEEDBACK WEIGHTING

The objective of the feedback weighting scheme is to identify a set of terms for query expansion. The refined query is assumed to reflect the user interactions more clearly. Hence, we apply a binary voting approach for term weighting and selection. BVM allows to weight terms and to rank them. In BVM, we can provide different weights to different implicit actions. The objective is to determine the influence of the use of implicit factors on retrieval results. Each implicit factor corresponds to a user interaction. The more interaction is accumulated on a shot the more important the shot will appear. We measure the importance in weighting the different factors. The different features identified in Section 3 are weighted for measuring the importance of a shot. If more actions appear on the same shot, the weighting of that shot should grow. Hence, we assume that a shot is more important when a user showed a higher interest in it. Implicitly detected shots can be used for query expansion which will hopefully improve retrieval results and help a user in retrieving more and better results. The advantage of implicit feedback is that a user does not explicitly have to define the relevance of a shot. Implicit feedback can improve results *without* disturbing the users' workflow. However, implicit factors can only provide an estimation of relevance. Giving explicit feedback, a user directly indicates whether a shot is relevant or not. Hence, explicit feedback is more reliable than implicit feedback and therefore should result in a higher weighting. Accordingly, implicitly detected results may not receive a higher weighting than explicitly selected ones.

Therefore, a user's implicit actions will manipulate the weighting of the shot and so, the weighting of its terms. Hopfgartner et al. [8] showed that different weighting factors for each implicit feature will influence retrieval results. They classified implicit user interactions into different categories. These categories can be weighted and cumulated, as a user might perform several of these interactions. The cumulated weighting can express the expanding relevance of a result. Based on that, we classified the user interactions and associated these implicit features we identified with a weighting presented in table 1. The weighting values varied between the different scenarios I_1–I_5.

Table 1: Weighting of Implicit Features

Action	W(I_1)	W(I_2)	W(I_3)	W(I_4)	W(I_5)
Highlighting	1.0	–	0.5	1.0	1.0
Click on a keyframe	0.5	1.0	1.0	1.0	1.0
Looking at metadata	–	–	–	–	1.0
Playing a video	(0–1)	(–)	(0–1)	(0–1)	(0–1)

Each feature will appear randomly in the simulation and their weighting will be combined and normalised to a "user feedback weighting".

Some feedback categories appear more often in a user interaction workflow than others, e.g. playing a video for a longer period of time. Therefore, we divided the playing duration into 0 – 10 time cycles, given a weighting of 0.1 for each cycle. Let's say, one cycle has a duration of 5 seconds. If we want to simulate the user playing a video for 10 seconds, we model it as playing a video for 2×5 seconds. This will result in the value 0.2.

A normalisation of the features will guarantee that the user feedback weighting will be between 0.0 and 1.0. This is important, as we are simulating explicit feedback in I_5. In our procedure, explicit feedback will give a user feedback weighting of 1 to the current shot.

The simulation of "browsing" or "using the sliding bar" does not increase the weighting of a shot. Instead, it has an influence on the list of shots which are taken for query expansion. If the "browsing" is simulated, the system adds the 0 – 10 right neighboured shots to the query expansion list. This will simulate a user browsing 0 – 10 times to the right neighboured shot. Based on our experience from our previous user study [14], we assume that a user mainly browses forward in time and rarely backwards. In the "sliding" simulation, we simulate a user jumping randomly 0 – 10 times

through the video. We take 0 – 10 random shots belonging to the same video and add them to the query expansion list.

An example simulating user behaviour I_1 (based on [8]) is as follows: We simulate the *highlighting of a tooltip*, *clicking on a keyframe* to trigger the video playback and *playing* that video for three time cycles. This simulated behaviour results in a normalised user feedback weighting of

$$\frac{1.0 + 0.5 + 0.3}{2.5} = 0.68$$

Another example simulating the user behaviour I_3 (based on [2]): We simulate the initial *click on a keyframe* to start *playing a video* for two time cycles. We do *not* simulate the highlighting of terms. Additionally, we simulate the *usage of the sliding bar* to randomly select three shots from the same video. This behaviour will reach a normalised user feedback weighting of

$$\frac{1.0 + 0.2}{2.5} = 0.48$$

and additionally, three random shots from the same video will be taken into account for the next query expansion, using the same weighting.

6. RETRIEVAL MODEL

In textual retrieval, miscellaneous models have been deployed to rank matching documents. One state-of-the-art approach in document retrieval is the probabilistic retrieval model BM25. This function ranks documents according to their relevance to a search query. In video retrieval, research has not been focused on the retrieval model. However, results from the textual domain promise relevance for the video domain. Nevertheless, the focus of this work is not on the retrieval model in video retrieval, but more on how to infer user needs from user behaviour. The aim is to improve adaptive multimedia retrieval interfaces. Thus, the BM25 retrieval model was used in this work.

7. EVALUATION METHODOLOGY

The classical Cranfield evaluation methodology in information retrieval employs test collections for the evaluation of retrieval engines. However, such a methodology is inadequate to evaluate interactive retrieval systems. Most interactive video retrieval systems are evaluated in laboratory based user experiments. There are many issues with such evaluation methodologies such as the lack of repeatability. In addition, to make a robust measurement, we need a large user population, which is very expensive. Besides, it is hardly possible to benchmark different combinations of features for effectiveness using user-centred evaluations.

An alternative way of evaluating such systems is the use of simulated interaction. In such an approach, we assume the possible steps a user may take if that person is sitting in front of the system. One such action is viewing relevant documents. The actions a serious user takes are expected to improve the retrieval of relevant documents. In an evaluation scheme called "simulated user evaluation" we assume some actions a real user may take and use them to influence further retrieval results. In this approach, we can benchmark different interactive retrieval approaches. The aim of this work is to explore the use of a simulated evaluation methodology to benchmark different interface schemes and also various implicit feedback schemes. In the next section, we will discuss our evaluation scheme.

8. SIMULATED EXPERIMENTS

To identify the best implicit indicators for relevance, we employed a methodology which simulates users giving implicit and explicit relevance feedback using the five different user behaviour models introduced in Section 4. The experiments make use of the TRECVID dataset which is presented in Section 8.1. The runs will be introduced in section 8.2. To perform test runs, we implemented a video retrieval system. Since the state-of-the-art video retrieval systems indicate better performances using textual components, we experiment within a text based video retrieval system. However, the same experiments can be performed with content-based features as well. The Terrier retrieval system [11], with the BM25 retrieval model, is used for indexing and retrieving based on textual components.

8.1 Data Collection

Our test runs are based on the 2006 TRECVID data set.[4] The set is approx. 160 hours of television news video in English, Arabic and Chinese language. The data set also includes the output of an automatic speech recognition system, the output of a machine translation system (Arabic and Chinese to English) and the master shot reference. A common collection of keyframes is also included. Each shot is considered as a separate document and is represented by text from the speech transcript. In the collection, we have

- 79484 number of shots
- 15.89 terms on average per shot
- 31583 shots without annotation

The data set contains search topics and relevance judgements. The search topics are designed to represent different types of queries real users might pose: request for video with specific types of people, specific instances of objects, specific activities or locations.

8.2 Simulation Methodology

8.2.1 Simulation Procedure

As explained above, we used the 24 topics/queries associated with the TRECVID data set. The relevant results associated with these topics (ground truth data) are given with the data set. Queries are given to the retrieval system and the results are produced using the Terrier retrieval [11] engine with the BM25 formulae. It is assumed that the user actions are aimed at retrieving relevant results. One such assumption is "looking through the relevant documents". Depending on the interface scenario used, a user may click on the keyframe, highlight a tool tip, look at metadata and/or play the video. The number of actions users perform on a result depend on random parameters. For each topic, we simulated user behaviour for a number of iterations.
The simulation process involves the following steps:

[4]The aim of the TRECVID workshop is to promote progress in content-based retrieval from digital video. The 2006 guidelines can be found online: http://www-nlpir.nist.gov/projects/tv2006/

- *Start retrieval.* The initial retrieval is triggered by a manually created query. The query for the next iteration is created at the end of each iteration (see last action of this list). Our retrieval will return a result list of shots.
- *Detecting top x relevant shots.* It is assumed that users select a number of relevant documents from the inital result list. The actual number of relevant documents persued depends on the scenario.
- *Expand queries from relevant shots.* We used the identified documents to expand query terms. The terms are fed into the system for retrieving a new set of documents.
- *Create weighting factor based on simulated user feedback.* We simulate the number of actions a user performs on results using the systems I_1 – I_5. As explained, the actions are based on random parameters. Using e.g. I_1, a possible action combination could be "tooltip highlighting" and "no click on a keyframe". Each action combination will raise a weighting based on the feedback weighting introduced in Section 5.
- *Combine weighting factor and expanded query terms.* Both weighting factor and the expanded query terms are combined and stored in a global query expansion list. If a term already exists in the list, the stored weighting and the new weighting will be combined. An example: The term-weighting combination "bush":"20" is already in the list. Now, the combination "bush":"10" shall be added. It will be updated in the expansion list as "bush":"30". This guarantees that frequent terms will receive a higher weighting in each iteration.
- *Create new query using top y weighted terms.* A new query is formed consisting of the top y weighted terms.

8.3 Results

As explained above, we need to select a number of parameters for the simulation. One is the initial query we used to start our simulations. Others are the number of detected relevant documents x, the percentage of relevant vs. non-relevant results and finally the number of terms y used for query expansion.

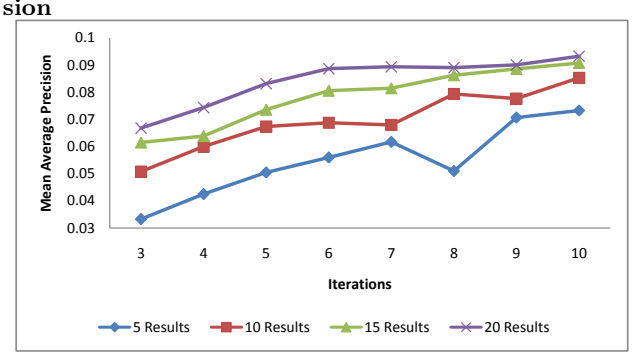
8.3.1 Initial Query

For each simulation, an initial query was given to the retrieval engine based on the search topic. They are manually created, based on the topic description. The reason for generating a set of terms from the topic manually is to create a good set of initial results. This is important for the simulation. A query consists of one to five terms with a median of 2 and an average of 2.5. The first retrieval returns an average of 12.9 (median: 9.5) relevant shots out of 100 results over all search tasks.

8.3.2 Number of Results

We compared simulation runs detecting the top five, top ten, top 15 and top 20 top relevant results x respectively (see Figure 6), identified by comparing the results with the provided ground truth data of that search topic. The more relevant shots are added to the query expansion list, the better the mean average precision. However, the more shots

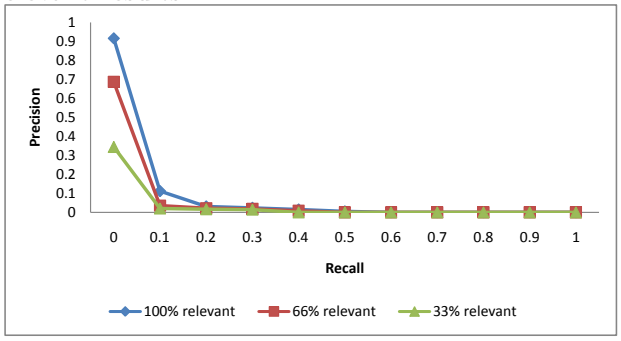
Figure 6: Number of results taken for query expansion



are taken into account, the smaller the improvement, compared to runs adding less shots. This is derives to the structure of the data set: The shots are associated with only a few keywords (15.89 terms on average per shot including stop/words), hence expanding more results will not result in many new terms. We can conclude from this that more relevant shots will return better results. Nevertheless, as the improvement steps get smaller the more results are taken into account, it might be better to perform a query expansion on a smaller set of results, as a user should receive new terms from query expansion rather early than later in the interaction process. In our simulation runs, we take the top five results into account. An average of 4.5 relevant shots (median: 5) can be found within the top five search results.

8.3.3 Relevant vs. Non-relevant Results

Figure 7: Precision/Recall of runs with x percent relevant results



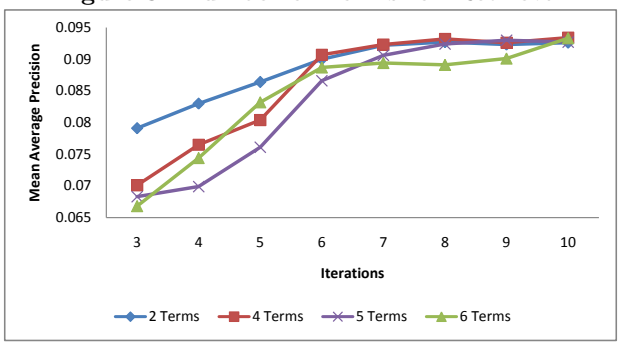
We also considered the simulation in which some of the non-relevant documents x were added to the query expansion list. Figure 7 illustrates Precision/Recall of each simulation run. The higher the number of non-relevant results taken for query expansion, the worse were the later retrieval results. The reason is obvious: A query expansion from terms of non-relevant results will reduce the percentage of relevant terms over each iteration. As our focus is on comparing different user scenarios, we only take relevant results taken from ground truth into account. This guarantees the best possible results in later iterations. Hence, we simulate a user clicking only on those results which appear to be

relevant. This is necessary as otherwise, our system will return too many non relevant results due to the already weak bounding between key terms and relevance in the TRECVID collection.

8.3.4 *Number of Query Terms*

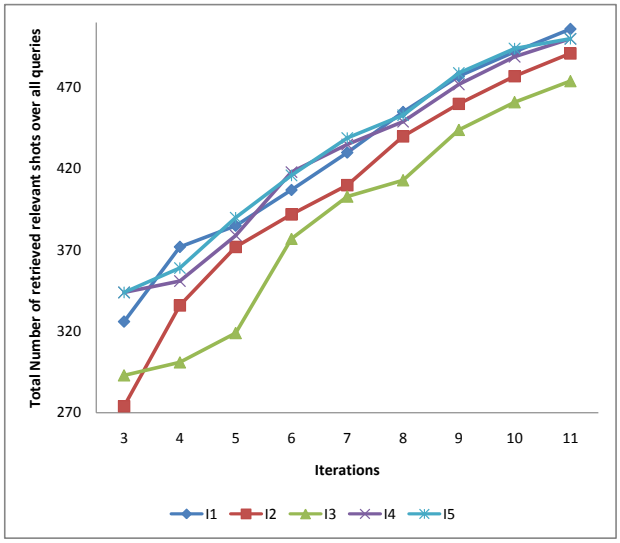
In the course of our research, we compared the mean average precision of retrieval runs using different numbers of terms y for retrieval. Results are illustrated in Figure 8. The more terms are taken to formulate a new query, the worse becomes the mean average precision during the first iterations. The reason is that fewer terms are more precise and set a stricter focus. The difference of Precision/Recall is minimal. Thus, less terms will return better retrieval results as they are more focused than more terms. In our simulations, we use a maximum of six terms, the top six terms that were detected so far.

Figure 8: Number of Terms for Retrieval



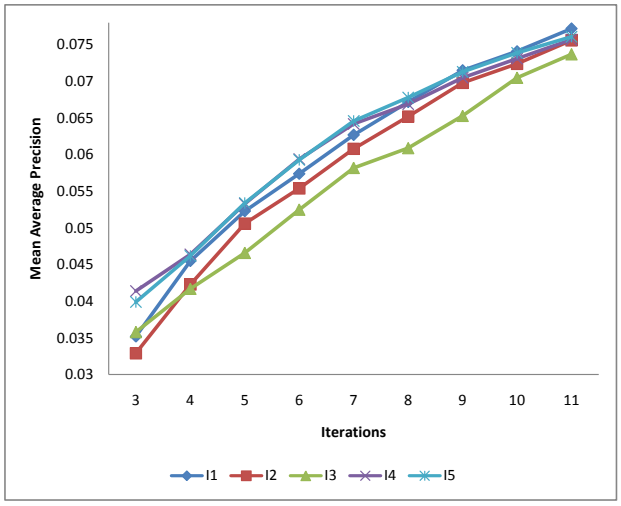
8.3.5 *Interfaces*

Figure 9: Total number of retrieved relevant shots over all queries



Figures 9 and 10 illustrate the results of the simulated tests. Figure 9 displays the total number of retrieved relevant shots over all queries over the relevance feedback iterations for the scenarios I_1 – I_5. As illustrated, the models I_1, I_4 and I_5 tend to return higher numbers of retrieved relevant shots over all queries than the other two models. Looking at the mean average precision of the test runs (see Figure 10), again I_1, I_4 and I_5 are the most successful models. Comparing both figures, I_3 shows the weakest performance.

Figure 10: Mean Average Precision



9. DISCUSSION

The aim of this study was to explore whether implicit indicators can be used to improve interactive video retrieval. Each scenario returned retrieval results that differ from the other scenarios. As each scenario is the simulation of implicit feedback given by a user, one can conclude that implicit indicators influenced the retrieval runs. The scenarios I_2 and I_3 include only few implicit features. As their results return the weakest retrieval results, it may hint to the assumption that using more implicit indicators can improve retrieval cycles.

One of the most significant results of our simulation is the similar performance of the systems I_1 and I_4. I_1 is our basic system while I_4 models the system of Hopfgartner et al. [8]. The only difference between them is that I_4 simulates the browsing through a video. This may indicate, that browsing can boost relevant retrieval results. This assumption is supported by the performance of I_5. It was the most successful model and also includes the simulation of browsing. Thus, I_5 was the only model which included the additional simulation of explicit relevance feedback. This correlates with the conclusions taken in the textual domain that the combination of explicit and implicit relevance feedback improves retrieval results.

10. CONCLUSIONS AND FUTURE WORK

Our work was focused on two aspects. One was developing an adaptive video retrieval strategy by making use of implicit features to improve retrieval results. We assume that both explicit and implicit indicators can improve retrieval in the video domain. To support this hypothesis, we analysed the influence of implicit features as an indicator for

relevance. Based on the interfaces of state-of-the-art adaptive video retrieval systems and the analysis of a small user study, we identified six implicit relevance features. We designed five different user interface scenarios I_1 – I_5 (based on state-of-the-art interface designs) which include different combinations of these six relevance features. Based on these scenarios, we ran a simulated user study to see, if the different combinations of features can have an influence on retrieval results. The results of our simulation are presented in Section 8.3 and are discussed in Section 9. They illustrate different performances for each user interface scenario. As the various user behaviour scenario simulations perform differently, we conclude from our work that implicit features *do* have an influence on interactive video retrieval results. The scenario I_5 performed best. This matches experiences from the textual retrieval domain as it was the only system that included the simulation of explicit relevance feedback.

This simulated methodology is a pre-implementation method. Given the numerous combinations of features and interface scenarios, we select an appropriate number of them. This will give a further opportunity to develop appropriate systems and subsequent user-centred evaluation. The real effect of a video retrieval system only can be measured by user experiments. The presented approach, however, provide a mechanism to benchmark a number of possible models before it reaches implementation.

11. REFERENCES

[1] P. Aigrain, H. Zhang, and D. Petkovic. Content-based representation and retrieval of visual media: A state-of-the-art review. *Multimedia Tools and Applications*, 3:179–202, 1996.

[2] P. Browne, C. Czirjek, G. Gaughan, C. Gurrin, G. Jones, H. L. S. Marlow, K. M. Donald, N. Murphy, N. O'Connor, N. O'Hare, A. F. Smeaton, , and J. Ye. Dublin City University Video Track Experiments for TREC 2003. In *TRECVID 2003 - Text REtrieval Conference TRECVID Workshop*, MD, USA, 2003. National Institute of Standards and Technology.

[3] M. Christel and R. Concescu. Addressing the Challenge of Visual Information Access from Digital Image and Video Libraries. In *Proc. ACM/IEEE-CS Joint Conference on Digital Libraries (Denver, CO, June 2005)*, pages 69–78, 2005.

[4] M. Claypool, P. Le, M. Wased, and D. Brown. Implicit interest indicators. In *Intelligent User Interfaces*, pages 33–40, 2001.

[5] E. Foley, C. Gurrin, G. Jones, C. Gurrin, G. Jones, H. Lee, S. McGivney, N. E. O'Connor, S. Sav, A. F. Smeaton, and P. Wilkins. TRECVid 2005 Experiments at Dublin City University. In *TRECVid 2005 – Text REtrieval Conference, TRECVID Workshop, Gaithersburg, Maryland, 14-15 November 2005*, 2005.

[6] M. Hancock-Beaulieu and S. Walker. An evaluation of automatic query expansion in an online library catalogue. *J. Doc.*, 48(4):406–421, 1992.

[7] D. Heesch, P. Howarth, J. Magalhães, A. May, M. Pickering, A. Yavlinski, and S. Rüger. Video Retrieval using Search and Browsing. In *TREC2004 – Text REtrieval Conference, Gaithersburg, Maryland, 15-19 November 2004*, 2004.

[8] F. Hopfgartner, J. Urban, R. Villa, and J. Jose. Simulated Testing of an Adaptive Multimedia Information Retrieval System. In *Proceedings of the Fifth International Workshop on Content-Based Multimedia Indexing (CBMI 2007), Bordeaux, France*, pages 328–335, 2007.

[9] R. Jesus, J. Magalhães, A. Yavlinski, and S. Rüger. Imperial College at TRECVID. In *TRECVid 2005 – Text REtrieval Conference, TRECVID Workshop, Gaithersburg, Maryland, 14-15 November 2005*, 2005.

[10] D. Kelly and J. Teevan. Implicit Feedback for Inferring User Preference: A Bibliography. *SIGIR Forum*, 32(2), 2003.

[11] I. Ounis, G. Amati, V. Plachouras, B. He, C. Macdonald, and D. Johnson. Terrier Information Retrieval Platform. In *Proceedings of the 27th European Conference on Information Retrieval (ECIR 05), Santiago de Compostela, Spain*, 2005.

[12] G. Salton and C. Buckley. Improving retrieval performance by relevance feedback. *Readings in Information Retrieval*, pages 355–364, 1997.

[13] A. Spink, H. Greisdorf, and J. Bateman. From highly relevant to not relevant: examining different regions of relevance. *Inf. Process. Manage.*, 34(5):599–621, 1998.

[14] J. Urban, X. Hilaire, F. Hopfgartner, R. Villa, J. Jose, C. Siripinyo, and Y. Gotoh. Glasgow University at TRECVID 2006. In *TRECVid 2006 – Text REtrieval Conference, TRECVID Workshop, Gaithersburg, Maryland, 13-14 November 2006*, 2006.

[15] R. White, J. Jose, C. van Rijsbergen, and I. Ruthven. A Simulated Study of Implicit Feedback Models. In *Proceedings of the 26th European Conference on Information Retrieval Research (ECIR '04). Lecture Notes in Computer Science*, 2004.

[16] R. W. White, J. Jose, and I. Ruthven. Adapting to Evolving Needs: Evaluating a Behaviour-Based Search Interface. In *Proceedings of 17th Annual HCI Conference (2nd Volume) Bath, UK, 2003*, 2003.