ABSTRACT
In this project (VIRSI) we investigate the promising content-based retrieval paradigm known as interactive search or relevance feedback, and aim to extend it through the use of synthetic imagery. In relevance feedback methods, the user himself is a key factor in the search process as he provides positive and negative feedback on the results, which the system uses to iteratively improve the set of candidate results. In our approach we closely integrate the generation of synthetic imagery in the relevance feedback process through a new fundamental paradigm: Artificial Imagination (AIM).

Categories and Subject Descriptors
H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – Query formulation, Relevance feedback, Search process.

General Terms
Algorithms, Human Factors.

Keywords
Content-based retrieval, principal component analysis, artificial imagination

1. INTRODUCTION
Beyond the borders of science, culture, and art, the issue of finding multimedia information has become one of the grand challenges of our time. There has been significant success in searching through text databases or multimedia using text annotations, but in many situations text annotation is not available for multimedia libraries (e.g. [14]). In such cases it is necessary to use content-based retrieval methods which directly analyze the pictorial content of the media.

One of the main areas explored in this project is the way in which artificial imagination can be used to aid a computer algorithm in learning new visual concepts, in particular to obtain meaningful solutions to user queries on multimedia databases.

Our own visual imagination allows us to create new and useful examples based on our memories and experiences. In artificial imagination we intend to harness similar strategies to intelligently synthesize informative examples. In particular, we intend to endow the computer with the ability to ask whether a particular synthesized example (which is not yet in the database) is relevant.

Such an example would for instance target a particular feature that could be important to the query. One example from real life is when a police officer creates a sketch of an unknown person to help identify a criminal.

2. BACKGROUND
The earliest years of Visual Information Retrieval (VIR) were characterized by "one-shot" methods based on the assumption that a single query would suffice to obtain useful results. Influential and popular examples of these systems would be the similarity-based search systems of QBIC ([6]) and Virage ([5]) circa mid 90s.

Beyond the one-shot queries in the early similarity-based search systems, the next generation of systems attempted to integrate continuous feedback from the user in order to gain improved insight into the user query. The resulting relevance feedback methods (e.g. [3, 4, 8, 9, 10, 11, 15, 21]) usually share the following process: the system shows the user a set of candidate results; the user labels a subset of the candidates as positive or negative examples; based on these examples the system reformulates the representation of the user query, and subsequently presents the user with the next set of candidate results. This process is repeated until, hopefully, the user is satisfied. Relevance feedback can be considered a special case of emergent semantics. Other names used in the computer vision literature have included query refinement, interactive search, and active learning.

A general approach is to view relevance feedback as a type of pattern classification where a relevant class is learned from a set of training examples with relevant and irrelevant labels. In principle, it is therefore possible to apply any type of learning algorithm in the relevance feedback loop.

One of the major problems in relevance feedback is how to address the typically very small training sets (e.g. [21]); in [20] it was found that combining multiple relevance feedback strategies gives superior results as opposed to any single strategy. In [17], Tieu and Viola proposed a method for applying the AdaBoost learning algorithm and noted that it is quite suitable for relevance feedback due to the fact that AdaBoost works well with small training sets. In [7] a comparison was performed between AdaBoost and SVM and found that SVM gives superior retrieval results. Good overviews can also be found in [12] and [21].

As in our system we mainly want to focus on the feedback-based generation of examples, we use the classic and well-known relevance feedback method proposed by Rocchio ([13]), where the simple idea is to move a query point toward the relevant examples and away from the irrelevant examples. The Rocchio
algorithm has the advantage of working relatively well when few examples are available. However, one challenging limitation of the Rocchio algorithm is that the single query point can necessarily refer to only a single cluster of results.

3. ARTIFICIAL IMAGINATION

Our visual imagination allows us to create newly synthesized examples based on our memories and experiences. When we are learning new visual concepts, we often construct such examples based on real objects or scenes to help understand or clarify the primary features which are associated with the concept. As mentioned, in this project we aim to explore the way in which artificial imagination by intelligently computed examples can be used to aid a computer algorithm in learning new visual concepts, and in particular to obtain meaningful solutions to user queries on multimedia databases.

In this context, artificial imagination can be considered as an advanced type of active learning: the systems asks the user specifically for feedback on a particular example in order to meet certain informational needs as well as possible; in the artificial imagination paradigm, examples are not restricted to be taken from the database itself like in traditional active learning, but are rather synthesized to more directly satisfy the informational requirements.

We thus understand “artificial imagination” here as the intelligent synthesis of examples based on known database entries and their location in the feature space. Potentially there are many techniques toward computing such examples. For example, we could use the current set of positive and negative examples to create a relevance space (RS): a space which is the same as the feature space but where each point in the feature space is attributed with an estimate of its relevance. A natural approach is then to examine the relevance space in relation to the locations of database items and attempt to clarify the relevance structure so that it affects the greatest number of database items.

Several novel challenges exist in the artificial imagination paradigm. Most importantly, we need to have methods to synthesize a virtual example based on a point in relevance space. The synthesis problem can be approached on a statistical level or directly by using transformations such as the Karhunen-Loeve Transform (KLT, e.g. [18]). In the case of the KLT, the feature space would be created by taking pairs of positive examples. In this case mixtures of these eigenvectors give a direct means of texture synthesis: given a feature vector \( c \) of coefficients representing the eigen-texture weights, a new image \( x \) may be synthesized through \( x = Qc + \mu \), where the columns of \( Q \) represent the eigen-textures, and \( \mu \) is an average image used for de-normalization.

As discussed earlier, the KLT-based decomposition also gives us a direct means of texture synthesis: given a feature vector \( c \) of coefficients representing the eigen-texture weights, a new image \( x \) may be synthesized through \( x = Qc + \mu \), where the columns of \( Q \) represent the eigen-textures, and \( \mu \) is an average image used for de-normalization.

As a first case-study we have designed a system for the retrieval of color texture images. The images are represented by means of a decomposition in terms of “eigen-textures” obtained through the KLT transform. Eigen-textures and image features have been determined for both entire texture images with and without color, as well as for various types of sub-blocks (both disjoint and overlapping).

Based on the positive and negative examples resulting from the most recent relevance feedback iteration, the Rocchio relevance ranking is determined by sorting the images based on their distance to the moving query point \( q_t \) given by

\[
q_t = \alpha q_{t-1} + \beta P_t - \gamma N_t, \tag{1}
\]

where \( P_t \) and \( N_t \) are the (possibly weighted) averages of the positive and negative example coefficients, respectively. The \( \alpha \), \( \beta \), and \( \gamma \) parameters tune the tradeoff between the attraction and repulsion of positive and negative examples, as well as the influence of past examples through \( q_{t-1} \).

We extend the Rocchio method to generate synthetic images based directly on expression (1). First of all, as the query point \( q_t \) generally does not correspond to an image in the database, we may thus use it to synthesize a new image. Further images may be generated by computing additional “Rocchio points” based on sub-samples of the feedback data. Using (1), each sub-sample of positive, and possibly negative, examples provides us with a feature vector that in turn may be transformed into a synthesized image. An extreme, yet useful, type of such subsets is obtained by taking pairs of positive examples. In this case mixtures of these positive examples, possibly offset to take into account previous feedback, are generated for use in further relevance feedback. If required, additional images may also be synthesized by varying the Rocchio parameters.

A natural line of thought is to indicate several images as relevant if a certain combination is expected to lead to a set of desired images. To illustrate, one may want to find regular crosshatched images by marking images that contain horizontal and vertical stripes as relevant, see Figure 1. Using the proposed synthesis method several synthetic images are then created, see Figure 2, and are presented to the user to solicit feedback. The user can then clarify that he is indeed looking for regular crosshatched images by selecting the second image, while a user that is looking for images containing horizontal or vertical lines, but not both, may want to select the third and fourth images.

4. RETRIEVAL SYSTEM

In our system we demonstrate how the performance of relevance feedback may be enhanced through the use of synthetically created images. Specifically, the system aims to offer proof of concept of such feedback by extending the classic Rocchio method to provide feedback-based artificially generated images. We also show how such images may naturally join in the feedback process, thereby providing a system where the image synthesis is fully integral to the relevance feedback process.

As a first case-study we have designed a system for the retrieval of color texture images. The images are represented by means of a decomposition in terms of “eigen-textures” obtained through the KLT transform. Eigen-textures and image features have been determined for both entire texture images with and without color, as well as for various types of sub-blocks (both disjoint and overlapping).

As discussed earlier, the KLT-based decomposition also gives us a direct means of texture synthesis: given a feature vector \( c \) of coefficients representing the eigen-texture weights, a new image \( x \) may be synthesized through \( x = Qc + \mu \), where the columns of \( Q \) represent the eigen-textures, and \( \mu \) is an average image used for de-normalization.

Based on the positive and negative examples resulting from the most recent relevance feedback iteration, the Rocchio relevance ranking is determined by sorting the images based on their distance to the moving query point \( q_t \) given by

\[
q_t = \alpha q_{t-1} + \beta P_t - \gamma N_t, \tag{1}
\]

where \( P_t \) and \( N_t \) are the (possibly weighted) averages of the positive and negative example coefficients, respectively. The \( \alpha \), \( \beta \), and \( \gamma \) parameters tune the tradeoff between the attraction and repulsion of positive and negative examples, as well as the influence of past examples through \( q_{t-1} \).

We extend the Rocchio method to generate synthetic images based directly on expression (1). First of all, as the query point \( q_t \) generally does not correspond to an image in the database, we may thus use it to synthesize a new image. Further images may be generated by computing additional “Rocchio points” based on sub-samples of the feedback data. Using (1), each sub-sample of positive, and possibly negative, examples provides us with a feature vector that in turn may be transformed into a synthesized image. An extreme, yet useful, type of such subsets is obtained by taking pairs of positive examples. In this case mixtures of these positive examples, possibly offset to take into account previous feedback, are generated for use in further relevance feedback. If required, additional images may also be synthesized by varying the Rocchio parameters.

A natural line of thought is to indicate several images as relevant if a certain combination is expected to lead to a set of desired images. To illustrate, one may want to find regular crosshatched images by marking images that contain horizontal and vertical stripes as relevant, see Figure 1. Using the proposed synthesis method several synthetic images are then created, see Figure 2, and are presented to the user to solicit feedback. The user can then clarify that he is indeed looking for regular crosshatched images by selecting the second image, while a user that is looking for images containing horizontal or vertical lines, but not both, may want to select the third and fourth images.

Figure 1. Regular crosshatch search: relevant image examples
Typically we observe small improvements in the relevance rankings after incorporating one or a small number of generated synthetic images as positive relevance feedback. As an illustration we show two image query sequences aimed at finding red images containing a pattern consisting of horizontal stripes. In the first sequence no synthetic images are used in the process, while in the second sequence they are. The initial set of images, including those selected as relevant, is shown in Figure 3. Normally, random images will be shown when starting a new query, but for illustration purposes we have used the same set of initial images for both sequences.

After two iterations, the first feedback sequence contains three images that are both red and predominantly contain the desired horizontal stripes, see Figure 4. In the second sequence a synthetic image that is generated from the query point is used as additional positive feedback, which results in a somewhat improved ranking, see Figure 5. Although the incorporation of synthetic images seems to only have modest positive effects, their use is nonetheless promising since the synthesized examples tend to show meaningful similarities with the positive examples, whereas the Rocchio query point often has a large distance to individual positive examples and sometimes centers on an undesirable cluster of images that show none or few of the desired image characteristics. In this case the synthetic images may offer valuable examples to steer the search to more relevant regions.

5. CONCLUSION AND FUTURE WORK
At the time of writing, we think this is the first system which integrates artificial imagination via image synthesis directly into the Rocchio-based relevance feedback process. In future work we
intend to explore more advanced relevance feedback strategies such as [19] as well as a method, presented in [10], for dealing explicitly with the partial relevance of specific image aspects.

The main future challenges will lie in the design of meaningful methods of image synthesis. To this end we will also explore evolutionary search strategies, such as presented in [1], [2] and [16]. For additional information, please refer to the URL http://www.liacs.nl/~bthomee/virsi/.

6. ACKNOWLEDGEMENTS

Our thanks to Bricks/NWO and LIACS at Leiden University for funding our research.

7. REFERENCES


