An Artificial Imagination for Interactive Search

Bart Thomee Mark J. Huiskes Erwin Bakker Michael S. Lew

LIACS Media Lab, Leiden University

{bthomee, Mark.Huiskes, erwin, mlew}@liacs.nl

Abstract. In this paper we take a look at the predominant form of human computer interaction as used in image retrieval, called interactive search, and discuss a new approach called *artificial imagination*. This approach addresses two of the grand challenges in this field as identified by the research community: reducing the amount of iterations before the user is satisfied and the small sample problem. Artificial imagination will deepen the level of interaction with the user by giving the computer the ability to think along by synthesizing ('imagining') example images that ideally match all or parts of the picture the user has in mind. We discuss two methods of how to synthesize new images, of which the *evolutionary synthesis* approach receives our main focus.

Keywords: Human computer interaction · Content-based image retrieval · Interactive search · Relevance feedback · Artificial imagination · Synthetic imagery · Evolutionary algorithms

1 Introduction

In the early years of image retrieval – the mid '90s – searches were generally performed using only a single query [3]. It was soon realized that the results could be significantly improved by applying interactive search and hence it did not take long before interactive search methods [4-9,11] were introduced into the field of image retrieval – or multimedia retrieval in general.

Interactive search, also known as relevance feedback, was initially developed to improve document retrieval [1]. Under this paradigm, the retrieval system presents a ranked set of objects relevant to the user's initial query and from thereon iteratively solicits the user for feedback on the quality of these objects and uses the feedback to compose an improved set of results. But, as keywords in a query can be matched one-on-one with text in a document and therefore good results can be obtained in a single step, its use has remained limited. In image retrieval, however, a user's query cannot be directly mapped onto items in the database. Interactive search turned out to be especially well suited for this problem: by interacting with the user, the system can learn which (features, parts of) images the user is interested in; the feedback resolves many of the uncertainties that arise as the system tries to learn what the user is looking for.

Despite the progress made, finding images of interest remains a major problem. Recent literature (e.g. [9], [11]), regards the following issues as the grand challenges in this field:

1. *Bridging the semantic gap* through improved concept detection techniques. Since users think in terms of high-level semantic concepts and not in low-level image features as available to the system, it is very important to select the most useful image descriptors, to help narrow this so-called semantic gap.
2. *Overcoming the curse of dimensionality* by selecting only the most optimal features. It is also essential to use suitable multi-dimensional indexing techniques for an efficient search in high-dimensional feature space, especially considering that performance quickly suffers with an increase in dimensionality [12].
3. *Reducing the amount of iterations* before the user is satisfied. Requiring a minimal amount of effort a user has to invest is key in relevance feedback: if too much involvement is demanded, the user will be reluctant to use the system.
4. *Solving the small sample problem*, which is the issue that the user will only label a few images while the amount of dimensions in feature space is huge, making it very difficult to discover the user's interest.

Note that solving (or at least alleviating) the small sample problem will have a direct effect on the amount of iterations needed: if more samples are labeled by the user, the retrieval system obtains better insight into the user's interests and consequently will be able to return better images, thus more quickly satisfying the user.

Several methods have been proposed that address the small sample problem. In [13] it was found that combining multiple relevance feedback strategies gives superior results as opposed to any single strategy. In [14], 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 [15] a comparison was performed between AdaBoost and SVM and found that SVM gives superior retrieval results. Good overviews can also be found in [16] and [11].

In section 2 we will address the third and fourth of above challenges by a new approach we call *artificial imagination*, which will deepen the level of interaction with the user by giving the computer the ability to think along by synthesizing ('imagining') example images that ideally match all or parts of the picture the user has in mind. In section 3 we will demonstrate our initial implementation of this approach.

2 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. One example from real life is when a journalist is looking for a photo to accompany his article and asks an archivist to find it, see Figure 1. *Artificial imagination* is the digital analogy of our own visual imagination. The computer is endowed with the ability to intelligently synthesize images and to present them to the user to ask whether or not they are relevant. These synthesized images are constructed in such a way that they target one or more particular features that are important to the query. Our idea is that the generated images (that are not in the database) are more in line with the user's thoughts and consequently the user will be able to select more images as relevant during an iteration. We can then compare the results obtained from search queries without using synthesized images to search queries including the synthesized images and see to what extent our approach improves the results.

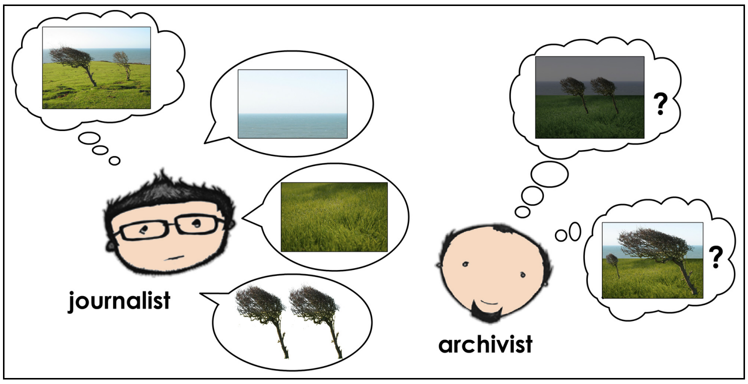

Fig. 1. An example of visual imagination. The journalist has a scene on its mind and tells the archivist what it looks like (sky, grass, trees). The archivist imagines scenes that contain the concepts mentioned; however because the image concepts might not be perfectly transferred, the imagined scenes do not necessarily initially resemble the image the journalist is thinking of (top-right imagined image). After exchanging image concepts ('night-time or day-time?', 'trees equal size or one larger than the other?') and obtaining more detailed information, the archivist is then better able to imagine the scene (bottom-right imagined image) and consequently is better able to return suitable images.

A content-based image retrieval system employs methods that analyze the pictorial content of the images in the database and performs similarity comparisons to determine which images the user most likely is interested in. Generally, the pictorial content is translated to a set of image features and, based on these, each image is then placed at the appropriate location in the high-dimensional feature space. The similarity comparisons are performed directly in this space (e.g. [1]) or after mapping this space to a lower (e.g. [17]) or higher (e.g. [18]) dimensionality.

As 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 [1], 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.

In Figure 2 a simplified system diagram is depicted of our content-based image retrieval system that uses synthetic imagery. It is important to realize that in our implementation the synthesis and retrieval/search aspects of the retrieval system are separate from each other: the feature space used to perform similarity comparisons does not necessarily have to be the same feature space used to synthesize images.

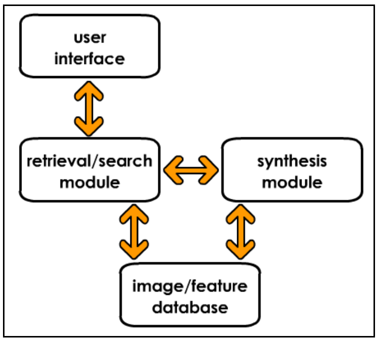

Fig. 2. Diagram of our retrieval system.

In our current system, we use several MPEG-7 [19] features (edge histogram, homogeneous texture, texture browsing and color layout) for the space in which we determine the similarity between images and in which we discover more relevant images by using the Rocchio method; for synthesis we use the feature space formed by taking the Karhunen-Loeve Transform (KLT, e.g. [20]) of each image in the database and using N coefficients from its KLT representation. Section 2.2 discusses in more detail how this is done. Thus, each image – real or synthesized – is associated with both an MPEG-7 and a KLT feature vector. This is illustrated in Figure 3.

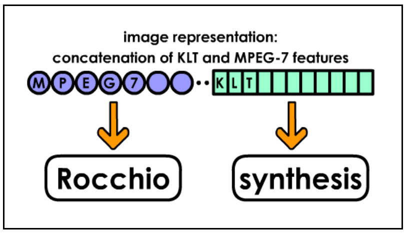

Fig. 3. The features used to represent an image serve different purposes.

The artificial imagination paradigm introduces several novel challenges, of which the most important ones are (i) which locations in feature space will be optimal candidates for synthesis and (ii) how do we synthesize an image given a point in feature space. In the following sections we suggest possible solutions to these two challenges.

2.1 Optimal synthesis locations

Here we describe two different methods to determine locations in the KLT feature space that are likely to result in suitable images when synthesized.

2.1.1 Inter-/extrapolation of feedback

By analyzing what effect the feedback has on the movement of the Rocchio query point over time, we can infer one or more locations where the query point likely will move after the next iteration; these points can therefore be synthesized, see Figure 4.

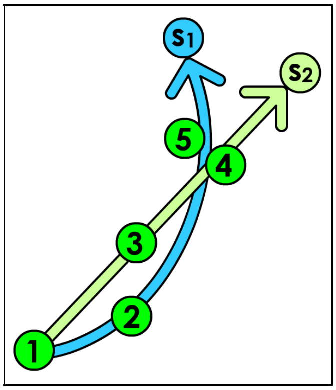

Fig. 4. Inferring likely future query points for synthesis: if over time the query point moved from point 1 and ending at 5, two likely future query points could be s_1 and s_2.

2.1.2 Evolution of feedback

Using evolutionary algorithms, points in feature space can be determined that are likely to maximize the 'fitness' (suitability, relevance) of the synthetic image. See Table 1 for the steps in our algorithm. After step 4 the algorithm loops back to step 2. An illustration of the crossover and mutation steps are shown in Figure 5.

Table 1. Evolutionary synthesis algorithm

step 1: **starting population**	random selection of images
step 2: **crossover**	sub-sampling: after feedback has been received, take subsets of the positive examples and mix their feature vectors to yield new points in feature space
step 3: **mutation**	negative examples and previous feedback can be used to move the points generated in step 2 to an adjusted location, or new random elements can be introduced into the syntesized images
step 4: **survival**	the user determines the fitness of the synthesized images by providing relevance feedback

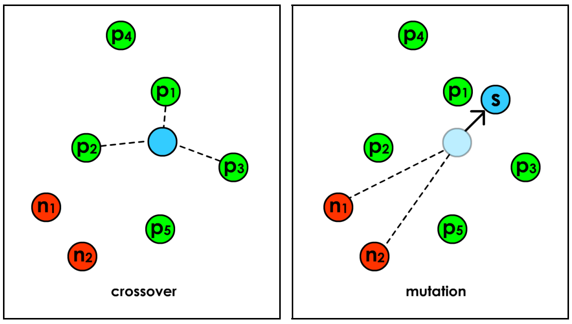

Fig. 5. Crossing over positive points p_1, p_2 and p_3 and mutating the resulting point through influence of negative points n_1 and n_2 to arrive at point s, from which a synthetic image can be generated.

2.2 Image synthesis

A point in KLT feature space is a set of coefficients weighting the eigenvectors of the KLT representation. We can thus synthesize the corresponding image by the standard method of linear reconstruction using the coefficients and corresponding eigenvectors. The MPEG-7 features of this synthesized image can be easily extracted by considering the image as any regular image and applying the MPEG-7 feature extractors to it, see Figure 6.

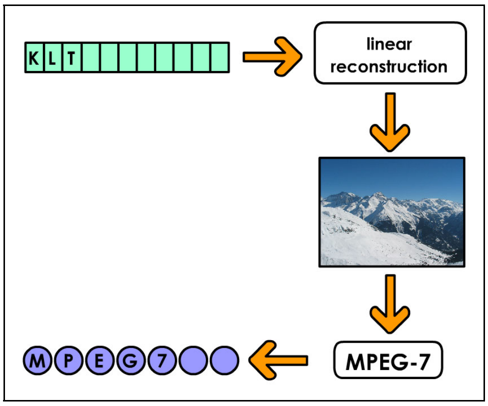

Fig. 6. From KLT feature point to synthetic image, followed by determining its MPEG-7 features. The KLT and MPEG-7 features together enable this new image to be used in the feedback process as if it is a real existing image from the image database.

For clarity, we give a detailed example in pseudocode of the process of finding similar images and subsequently synthesizing a few images with the evolutionary synthesis algorithm. In this example, the user only selects relevant (positive) images from the initial random selection.

```
'preprocessing
load all N images from disk into I[n], where n = 1...N
for each n, do
   compute MPEG-7 features of image I[n] and store in M[n]
   compute KLT coefficients of image I[n] and store in K[n]
end

'initial selection
select R random images from I; present to user

wait for user to perform relevance feedback
store positive examples in P[i], i = 1...NP

'analyze feedback
given relevant images P[i] do
   determine Rocchio query point q = mean(M[P[i]])
   select most similar images W to q using distance to M[n]
   'crossover
   for each nonempty subset Ak⊂P[i], k = 1...(2^NP-1) do
      'synthesize new feature vector
      K[N+k] = mean(K[P[Ak]])
      synthesize image I[N+k] by inverse KLT on K[N+k]
      compute MPEG-7 features M[N+k] of image I[N+k]
   end
   return W and S = (I[N+1]...I[N+2^NP-1]) to the user
end
```

One should realize that the synthesis of an image using a feature space other than KLT may not be as straightforward. With KLT, a direct one-on-one mapping exists between the coefficients/eigenvectors and the pixels through linear reconstruction. This is generally not the case with other feature spaces where image reconstruction is not well-defined. For instance, suppose that color histograms are used as features. Given a color histogram, it is not possible to synthesize a unique image. We have information about how often each color should appear in the image, but we do not know where the colors should be located, and many images have the same histogram.

3 Results and Examples

We have developed a system for the retrieval of color texture images which uses 1000 textures taken from the Corel database. The images are represented by means of a decomposition in terms of "eigen-textures" obtained through the KLT transform and are additionally associated with several MPEG-7 features (edge histogram, homogeneous texture, texture browsing and color layout). The synthetic images we show are created using our evolutionary algorithm approach.

The following example provides a first proof-of-concept: using our current implementation, we typically observe small improvements in the relevance rankings after incorporating one or a small number of generated synthetic images as positive relevance feedback. As illustration we describe two image query sequences aimed at finding flowers/leaves that contain purple and green. In the first sequence no synthetic images are used in the process, while in the second sequence they are. The initial set of images that we indicate as relevant are shown in Figure 7.

Fig. 7. Initial selection of relevant images.

The results that are returned after submitting this query are shown in Figure 8. When not using any synthetic images that are generated, the user selects the five relevant images (shown with the green borders in Figure 8), and submits this modified query. The result from this second query is that the system is unable to return any new relevant images.

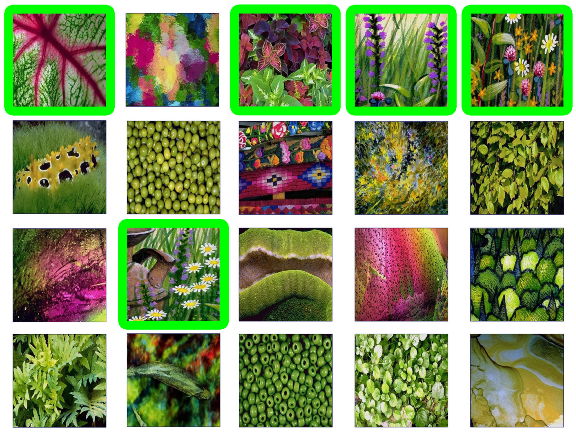

Fig. 8. Image ranking after the first step.

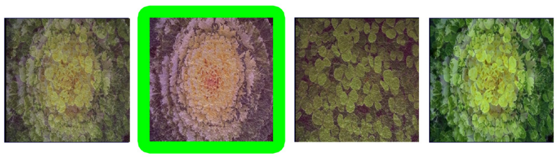

Fig. 9. Synthesized images by applying the evolutionary algorithm to the five relevant images that were selected in the initial screen.

However, if the user had included the most relevant image from the set of synthesized images – the purple flower with the green border in Figure 9 – then the results would have improved with an additional two relevant images, see Figure 10.

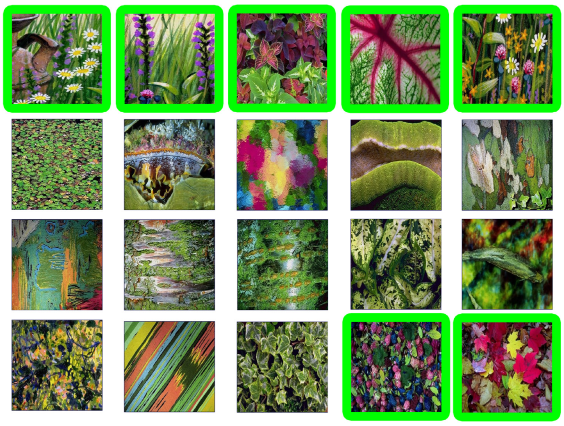

Fig. 10. Ranking after two iterations using the synthesized image.

Although the incorporation of synthetic images seems to only have modest positive effects, their use is nonetheless promising. 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.

4 Conclusions and Future Work

Artificial imagination is a promising paradigm that can significantly enhance the level of interaction of the retrieval system with the user. By giving the retrieval system the power to imagine, it will more quickly understand what the user is looking for. Our evolutionary synthesis algorithm shows much potential for the generation of synthetic imagery. In future work we intend to explore other classification/synthesis methods such as wavelets [10], more advanced relevance feedback strategies such as [18], methods for dealing explicitly with the partial relevance of specific image aspects [8], and other focused areas such as face analysis and recognition [2]. Also, to apply the concept of artificial imagination to the case of general images (as opposed to textures only), we will work on the generation of collages, where the system can combine image concepts/objects by placing them in a single image.

Acknowledgments

We would like to thank Leiden University, University of Amsterdam, University of Illinois at Urbana-Champaign, the Dutch National Science Foundation (NWO), and the BSIK/BRICKS research funding programs for their support of our work.

References

1. Rocchio, J.J.: Relevance Feedback in Information Retrieval. The Smart Retrieval System: Experiments in Automatic Document Processing. G. Salton, Ed. Prentice-Hall (1971)
2. Cohen, I., Sebe, N., Garg, A., Lew, M.S., Huang, T.S.: Facial expression recognition from video sequences. In: Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), Lausanne, Switzerland, vol. I, pp. 641-644 (2002)
3. Bach, J.R., Fuller, C., Gupta, A., Hampapur, A., Horowitz, B. Humphrey, R. Jain, R. Shu, C.-F: Virage Image Search Engine: An Open Framework for Image Management. In: Proceedings of the SPIE Storage and Retrieval for Still Image and Video Databases, p.p. 76-87 (1996)
4. Chang, S.-F., Chen, W., Sundaram, H.: Semantic Visual Templates: Linking Visual Features to Semantics. In: Proceedings of the IEEE International Conference on Image Processing, p.p. 531–535, IEEE Computer Society Press (1998)
5. Chen, Y., Zhou, X.S., Huang, T.S.: One-class SVM for Learning in Image Retrieval, In: Proceedings of IEEE International Conference on Image Processing, p.p. 815-818 (2001)
6. Haas, M., Rijsdam, J., Thomee, B, Lew, M.S.: Relevance Feedback: Perceptual Learning and Retrieval in Bio-computing, Photos, and Video, In: Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval, p.p. October, 151-156 (2004)
7. He, X., Ma, W.-Y., King, O., Li, M., Zhang, H.: Learning and Inferring a Semantic Space from User's Relevance Feedback for Image Retrieval. In: Proceedings of the ACM Multimedia. p.p. 343–346 (2002)
8. Huiskes, M.J.: Aspect-based Relevance Learning for Image Retrieval. In: Proceedings of the International Conference on Image and Video Retrieval, Singapore, Lecture Notes in Computer Science 3568, p.p. 639-649 (2005)
9. Lew, M.S., Sebe, N., Djeraba, C., Jain, R.: Content-based Multimedia Information Retrieval: State of the Art and Challenges. ACM Transactions on Multimedia Computing, Communications, and Applications 2(1), p.p. 1-19 (2006)
10. Sebe, N., Lew, M.S.: Wavelet Based Texture Classification. In: Proceedings of the International Conference on Pattern Recognition, vol. III, pp. 959-962 (2000)
11. Zhou, X.S., Huang, T.S.: Relevance Feedback in Image Retrieval: A Comprehensive Review, Multimedia Systems Journal 8(6), p.p. 536-544, Springer-Verlag (2003)
12. Böhm, C., Berchtold, S.: Searching in High-Dimensional Spaces: Index Structures for Improving the Performance of Multimedia Databases. In: ACM Computing Surveys 33(3), p.p. 322-373 (2001)
13. Yin, P.-Y., Bhanu, B., Chang, K.-C., Dong, A.: Integrating Relevance Feedback Techniques for Image Retrieval Using Reinforcement Learning. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(10), p.p. 1536- 1551 (2005)
14. Tieu, K., Viola, P.: Boosting Image Retrieval. International Journal of Computer Vision 56(1), p.p. 17-36 (2004)
15. Guo, G., Zhang, H.-J., Li, S.Z.: Boosting for Content-Based Audio Classification and Retrieval: An Evaluation. In: Proceedings of the IEEE Conference on Multimedia and Expo, (2001)
16. Muller, H., Muller, W., Marchand-Maillet, S., Pun, T., Squire, D.: Strategies for Positive and Negative Relevance Feedback in Image Retrieval. In: Proceedings of 15th International Conference on Pattern Recognition, p.p. 1043-1046 (2000)
17. Lin, Y.-Y., Liu, T.-L. Liu, Chen, H.-T.: Semantic manifold learning for image retrieval. In: Proceedings of the 13th annual ACM international conference on Multimedia, pp. 249-258 (2005)
18. Tong, S., Chang, E., Support Vector Machine Active Learning for Image Retrieval. In: Proceedings of the 9th ACM International Conference on Multimedia, p.p. 107-118 (2001)
19. Manjunath, B.S., Salembier P., Sikora, T.: Introduction to MPEG-7: Multimedia Content Description Interface, John Wiley & Sons (2002)
20. Therrien, C.: Decision, Estimation, and Classification. John Wiley & Sons (1989)