14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

IMPROVING CONTENT-BASED IMAGE RETRIEVAL BY MODELLING THE SEARCH PROCESS: A BAYESIAN APPROACH

Simon P. Wilson and Georgios Stefanou

Department of Statistics, Trinity College
Dublin 2, Ireland
phone: + (353) 1 608 1759, fax: + (353) 1 677 0711, email: simon.wilson@tcd.ie
web: http://www.tcd.ie/Statistics/staff/simonwilson.shtml

ABSTRACT

In this paper we look at a simple image retrieval with relevance feedback scenario where we model simple properties of the search process. A content-based image retrieval method based on Bayesian inference is proposed that infers these search properties, as well as providing relevant images, from relevance feedback data. The approach is evaluated by performing searches for categories of image that invoke different emotional reactions.

1. INTRODUCTION

In this paper we develop a Bayesian approach to content-based image retrieval (CBIR). This is a subject with a large and expanding literature [4]. The work described builds on the relevance feedback mechanism of the PicHunter Bayesian retrieval system of Cox et. al. [2]. The principal contributions are to extend the Bayesian inference implemented in that paper to other aspects of the retrieval process, and to use this extension to aid the search. In particular, we define a model for the search process that contains parameters that crudely describe how well the feature space is describing the query and how the user conducts the search. We investigate whether the inference is consistent between users and different queries. This is tested on queries for images in a database of paintings with different emotional content. Our tentative conclusion, based on a small evaluation with real users, is that there is evidence that the method can capture useful aspects of the search process that can aid the retrieval process.

The paper is organised as follows. In Section 2 we describe a model for the retrieval process with relevance feedback. Section 3 is an evaluation; we investigate whether information that we learn about the search process for a particular query improves the search process subsequently. Section 4 is some concluding remarks.

2. MODEL FOR THE RETRIEVAL PROCESS WITH RELEVANCE FEEDBACK

We consider a database of images $\mathscr{I} = \{T_1, \ldots, T_N\}$. The objective is to determine the target image $T \in \mathscr{I}$. We assume the retrieval is of the mental-image type; no query image is provided but the user has an idea of an image in his or her mind that is to be retrieved [1]. A set of N_D images from $\mathscr{I}$, from which the user picks the most relevant. Let $D_i \subseteq \mathscr{I}$ to be the set of displayed images at the ith iteration of this process, and $A_i \in D_i$ to be the image picked. We define $H_t = \{D_1, A_1, D_2, A_2, \ldots, D_t, A_t\}$ to be the history of displayed images and user actions up to the tth iteration.

The Bayesian approach requires that a likelihood be defined for data — in this case the relevance feedback — in terms of the unknown quantities of interest. The principal unknown is T, but we also seek to learn more about the search process. We propose two further parameters: a scaling parameter σ that is interpreted as measuring the ability of the feature space to model the query, and a feature subset indicator F. Generalising the work in [2], the likelihood is the probability that a user picks A_k from D_k given the target image T, feature subset F and a scale σ, which we define to be:

$$\mathbb{P}(A_k \,|\, D_k, T, \sigma, F) = \frac{\exp\left(-d_F(A_k, T)/\sigma\right)}{\sum_{T_j \in D_k} \exp\left(-d_F(T_j, T)/\sigma\right)}, A_k \in D_k, \tag{1}$$

where d_F is Euclidean distance measure in the subset F of normalised image features.

Note that as $\sigma \to 0$ then with probability 1 the image in D_k that is closest to T in feature space will be picked, whereas as $\sigma \to \infty$ then all images are equally likely to be picked by the user. In this sense, σ is a measure of how well the feature space models the user's query. In this work we partition the feature space vector into three sub-vectors: global colour, texture and segmentation (object location and size) features; thus $F \in \{GC, TX, SG\}$. This is similar to probabilistic feature relevance techniques; see [3].

After the tth iteration the posterior probability distribution of T given H_t is computed:

$$\mathbb{P}(T, \sigma, F \,|\, H_t) \propto \left(\prod_{k=1}^{t} \mathbb{P}(A_k \,|\, D_k, T, \sigma, F) \right) \mathbb{P}(T, \sigma, F), \tag{2}$$

where $\mathbb{P}(T, \sigma, F)$ is the prior distribution on T, F and σ. Here we assume independent uniform priors:

$$\mathbb{P}(T, \sigma, F) = \mathbb{P}(T)\,\mathbb{P}(\sigma)\,\mathbb{P}(F) = \frac{1}{N} \frac{1}{\sigma_{\max}} \frac{1}{3}, \tag{3}$$

for $T \in \mathscr{I}, 0 \le \sigma \le \sigma_{\max}, F \in \{GC, TX, SG\}$.

The marginal distribution of T is then

$$\mathbb{P}(T \,|\, H_t) = \sum_{F \in \{GC, TX, SG\}} \int \mathbb{P}(T, \sigma, F \,|\, H_t)\, d\sigma.$$

The next display set D_{t+1} is taken to be the N_D most probable images from $\mathbb{P}(T \,|\, H_t)$, is displayed. This process repeats until the user encounters a satisfactory image. This posterior distribution of T represents the system's state of knowledge about relevant images and has taken account of what we have learned about the performance of the feature space — through σ — and which sub-vectors are most important for the relevance feedback through F.

3. EVALUATION

The system is evaluated on a database of 1066 paintings from the Bridgeman Art Library. This database contains images with a wide variety of subjects and styles. The purpose of the evaluation is to assess if information captured about the retrieval process — quantified by σ and F — can aid the search. The metric is the number of iterations of the process that are required to find a relevant image. We use a challenging query: find a painting in the database with romantic content.

The evaluation is done by running the same query 10 times, each time using the uniform prior distribution. A different random set of images is used as D_1 each time. We denote the data from the mth run of the query as $H_t^{(m)} = (D_1^{(m)}, A_1^{(m)}, \ldots, D_{t^{(m)}}^{(m)})$. The data from these 10 runs are combined to produce a posterior distribution for (T, σ, F):

$$\begin{aligned}
&\mathbb{P}(T,\sigma,F \,|\, H_t^{(1)},\ldots,H_t^{(10)}) \\
&\quad\propto \left(\prod_{m=1}^{10} \mathbb{P}(H_t^{(m)} \,|\, T,\sigma,F)\right) \mathbb{P}(T,\sigma,F), \\
&\quad= \left(\prod_{m=1}^{10}\prod_{k=1}^{t^{(m)}} \mathbb{P}(A_k^{(m)} \,|\, D_k^{(m)}, T,\sigma,F)\right) \mathbb{P}(T,\sigma,F),
\end{aligned}$$

where $\mathbb{P}(T,\sigma,F)$ is the prior of Equation 3 with $\sigma_{\max} = 20$. The posterior distribution of σ and F is

$$\mathbb{P}(\sigma,F \,|\, H_t^{(1)},\ldots,H_t^{(10)}) = \sum_{T\in\mathscr{I}} \mathbb{P}(T,\sigma,F \,|\, H_t^{(1)},\ldots,H_t^{(10)}).$$

The query is then repeated another 10 times, but using the above posterior for σ and F as the prior. The prior for T is still assumed uniform, so the prior for stage 2 is:

$$\mathbb{P}(T,\sigma,F) = \frac{1}{N}\,\mathbb{P}(\sigma,F \,|\, H_t^{(1)},\ldots,H_t^{(10)}).$$

We do not use information about T from the first set of queries as the target that is found is very dependent on the initial random display set and we do not want to confound this effect with one due to information about σ and F. The second set of queries was done about 6 months after the first to reduce any effect due to learning to use the system better.

Tables 1 and 2 show the results from the two stages. We note that many of the same images were discovered both within and between the two stages. The mean and standard deviation in number of iterations to discover these images are given at the bottom of the table. It is clear that the mean number of iterations for the second stage is less, but the difference is not statistically significant; a 2-sample t-test to compare the means gives a value of 1.51.

Figure 1 shows the posterior of σ and F after the first stage and after the second stage (including data from both stages). We observe that global colour features are inferred to be dominant in searching for romantic images. The important feature of these two distributions is that they are consistent; the posterior after the second stage puts mass on the

Images	Iterations
	2
	6
	16
	4
	3
	21
	2
	5
	4
	11
Mean	7.4
Std. Dev.	6.5

Table 1: Results of the ten queries in the first stage, with a uniform prior on (T,σ,F). Shown are the found images and the number of iterations required to find them.

Images	Iterations
	2
	6
	3
	2
	2
	8
	4
	2
	7
	5
Mean	4.1
Std. Dev.	2.3

Table 2: Results of the ten queries in the second stage, with a uniform prior on T and a prior for (σ, F) that is the posterior from all the data of the first stage. Shown are the found images and the number of iterations required to find them.

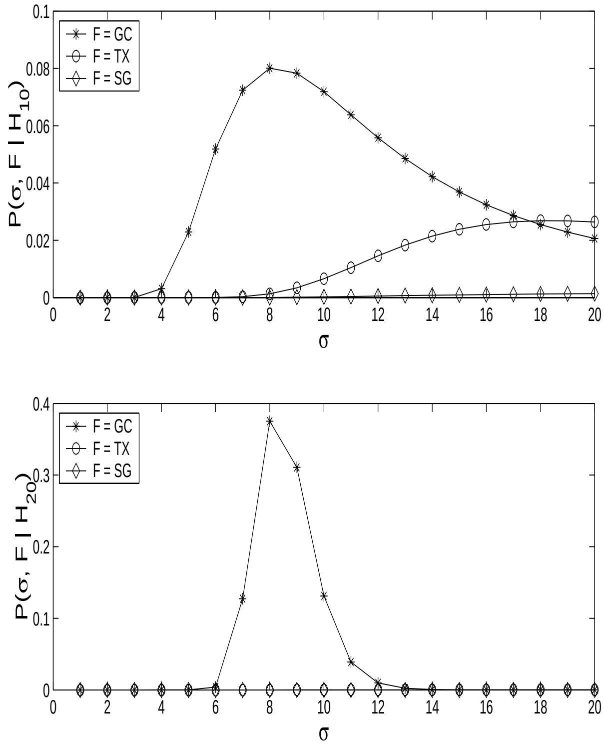

Figure 1: The posterior distribution of F and σ given the data for romantic image search after the first stage (top) and both stages (bottom).

same values as that after the first stage, although has smaller variance as it is conditional on more data.

Finally, we investigated if information on (σ, F) from one person is of use to another. A second person was asked to conduct the two stages of the evaluation, but this time the second stage prior was that from the first person (e.g. that used to obtain the results in Table 2). The mean and standard deviation of the number of iterations for the 10 queries of the first stage were 9.4 and 6.2 respectively. The mean and standard deviation of the number of iterations for the 10 second stage queries were 3.7 and 2.1 respectively. A 2-sample t-test statistic to compare the two is 2.75, which is statistically significant. So in this case we conclude that the mean number of iterations in the second stage was significantly less than the number in the first stage.

More details on these results can be found in [5].

4. CONCLUDING REMARKS

In this paper we have demonstrated a Bayesian approach to content-based image retrieval that tries to infer information about aspects of the search process. The tentative conclusion that we draw from Section 3 is that there is evidence of a decrease in the number of iterations to retrieve similar images. However, this is based on a small sample of 2 users. The information used from one run to another appears to be consistent, as evidenced by the consistency of the posterior distribution before and after the first stage in Figure 1. Further, there is some evidence that information from one user is of use to another for the same query.

Nevertheless this is a very small study and further experimentation, with other queries and other users, must be made before more general conclusions can be drawn.

ACKNOWLEDGEMENTS

The images of paintings are courtesy of the Bridgeman Art Library, London (`www.bridgeman.co.uk`). This work has been made possible by the Network of Excellence MUSCLE, contract number FP6-507752 (`www.muscle-noe.org`), funded by the European Union.

REFERENCES

[1] N. Boujemaa, J. Fauqueur, and V. Gouet. What's beyond query by example? In L. Shapiro, H. P. Kriegel, and R. Veltkamp, editors, *Trends and Advances in Content-Based Image and Video Retrieval*. Springer Verlag, 2004.

[2] I. J. Cox, M. L. Miller, T. P. Minka, T. V. Papathomas, and P. N. Yianilos. The Bayesian image retrieval system, PicHunter: theory, implementation and psychophysical experiments. *IEEE Transactions on Image Processing*, 9:20–37, 2000.

[3] J. Peng, B. Bhanu, and S. Qing. Probabilistic feature relevance learning for content-based image retrieval. *Computer Vision and Image Understanding*, 65:150–164, 1999.

[4] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.

[5] G. Stefanou. *Bayesian approaches to content-based image retrieval*. PhD thesis, School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland, 2005.