Score Distribution Approach to Automatic Kernel Selection for Image Retrieval Systems

Anca Doloc-Mihu and Vijay V. Raghavan

University of Louisiana at Lafayette, Lafayette, LA 70504, USA,
anca@louisiana.edu, raghavan@@cacs.louisiana.edu,
http://www.cacs.louisiana.edu/~axd9917

Abstract. This paper introduces a kernel selection method to automatically choose the best kernel type for a query by using the score distributions of the relevant and non-relevant images given by user as feedback. When applied to our data, the method selects the same best kernel (out of the 12 tried kernels) for a particular query as the kernel obtained from our extensive experimental results.

1 Introduction

This paper contributes to the design and evaluation of learning methods employed in Web-based Adaptive Image Retrieval System (AIRS) in order to maximize user benefits. Here, we focus on kernel-based learning methods. Specifically, the goal of this paper is to investigate the selection of the kernel type for a Web-based AIRS.

In a Web-based Image Retrieval System, the goal is to answer as best (fast and accurate) as possible to the user's request, given here as query image(s) represented by color histograms. However, a characteristic of these colors is that they are not independent, but correlated. More, the interaction between them is stronger for some queries (images) than for other queries.

In the former case, a more complex, possibly non-linear, kernel has to be used, whereas in the latter case, a linear kernel may be sufficient. Since we are dealing with real images, and therefore, with complex queries, it is expected that we need to use non-linear kernels to achieve good retrieval results [6]. However, in the literature, "there are currently no techniques available to learn the form of the kernel" [4].

Methods like Relevance Vector Machines [12] assume distribution of data, which might fit or not a real image collection. In our previous work [7], we used the score distribution models briefly presented in Section 2.1. Instead, in here, we propose a kernel selection method based on the relevance feedback information given by user. Our method computes the score distributions based on the information given by user as feedback (relevant and non-relevant images), without using any particular model for these distributions.

Based on the Kernel Rocchio [1] learning method, several kernels having polynomial and Gaussian Radial Basis Function (RBF) like forms (6 polynomials and 6 RBFs) are applied to general images represented by color histograms in RGB and HSV color spaces. We implement and test these kernels on two image collections of sizes 5000 and 10000. From the plots of these results, we select the best (effective and efficient) kernel desired to be used for each query.

Then, we applied our method to these collections for the same queries. For each query, the method selects the same best kernel (out of the 12 tried kernels) for a particular query as the kernel obtained from our extensive experimental results.

The paper organization is as follows. Section 2 presents a new method for automatically selecting the kernel to be used by the Kernel Rocchio learning method in order to improve the performance of an adaptive image retrieval system (AIRS). Section 3 reports experimental results. Finally, Section 4 concludes the paper.

2 Score Distribution Approach for Kernel Type Selection

Researchers [9, 10, 8] modeled the score distributions of search engines for relevant and non-relevant documents by using a normal and an exponential distribution, respectively. Based on this idea, in this section, we propose a procedure based on score distributions to automatically select the kernel type for an AIRS.

2.1 Mathematical Model of Image Score Distributions

The score distributions of the relevant documents suggested by researchers [10] are modeled as

$$P(score|\mathrm{P} = \mathrm{R}) = \frac{1}{\sqrt{2\pi\sigma}} \exp(-\frac{(\mathrm{score}-\mu)^2}{2\sigma^2}),$$

where μ is the mean, and σ is the variance of the Gaussian distribution. The score distributions of the non-relevant documents are modeled as

$$P(score|\mathrm{P} = \mathrm{NR}) = \lambda \exp(-\lambda * \mathrm{score}),$$

where λ is the mean for the exponential distribution.

Our experimental results (Section 3.1) support this mathematical models of the score distributions, with some approximation. Figure 1 illustrates how the score distributions fit the top 300 images for one query for 6 kernels.

As an observation, the score distributions are different for the different kernel types used in the experiments (Figure 1). This motivates us to seek for a method to select the kernel type by using these differences between the score distributions.

2.2 Kernel-based Retrieval

In Image Retrieval, the user searches a large collection of images, for instance, that are similar to a specified query. The search is based on the similarities of the image attributes (or features) such as colors. Then, a linear retrieval form [2, 1] matches image queries against the images from collection

$$F : R^N \times R^N \to R, F(\mathrm{P}, \mathrm{Q}) = \mathrm{P}^t\mathrm{Q},$$

where query image Q contains the features desired by user, and t stands for the transpose of vector P. The bigger the value of the function F applied to a query Q and an image P, the better the match between the query image Q and the collection image P, or, in other words, the closer the two images. For representing images by color, we use the histogram representations in RGB and HSV color spaces.

Recently, image retrieval systems start using different learning methods for improving the retrieval results. In this work, we use the **Kernel Rocchio** method [1] for learning, which combines the simplicity of Rocchio's method with the power of the kernel method to achieve improved results. For any image P_k from the collection, the algorithm computes its retrieval status values (RSV) as:

$$RSV(\phi(P_k)) = \frac{1}{|R|} \sum_{P_i \in R} \frac{K(P_i, P_k)}{\sqrt{K(P_i, P_i) \cdot K(P_k, P_k)}} - \quad (1)$$

$$\frac{1}{|\bar{R}|} \sum_{P_j \in \bar{R}} \frac{K(P_j, P_k)}{\sqrt{K(P_j, P_j) \cdot K(P_k, P_k)}}, \quad (2)$$

where R is the set of relevant images and $\bar{R}$ is the set of the non-relevant images.

The kernel method constitutes a very powerful tool for the construction of learning algorithm by providing a way for obtaining non-linear algorithms from algorithms previously restricted to handling only linearly separable datasets.

In this work, the general form of the polynomial kernels [4, 1] is given by

$$K(\mathrm{x}, \mathrm{y}) = (\langle \mathrm{x}, \mathrm{y} \rangle)^{\mathrm{d}}, \mathrm{d} > 0, \quad (3)$$

where $\langle \cdot, \cdot \rangle$ is the inner product between vectors x and y. The general form of the radial kernels is given by

$$K(\mathrm{x}, \mathrm{y}) = \exp\left(-\frac{\left(\sum_{\mathrm{i}=1}^{\mathrm{N}} |\mathrm{x}_\mathrm{i}^\mathrm{a} - \mathrm{y}_\mathrm{i}^\mathrm{a}|^\mathrm{b}\right)^\mathrm{c}}{2\sigma^2}\right), \sigma \in \mathrm{R}^+. \quad (4)$$

2.3 Kernel Selection Approach Based on Image Score Distributions

In image retrieval, the goal is to retrieve as many relevant images and as few non-relevant images as possible. This means, we wish to have a retrieval system capable to rank all the relevant images before the non-relevant ones. In other words, the scores of the relevant images should be as higher as possible, whereas the scores of the non-relevant images should be as lower as possible.

In our system, these scores are computed by using the different kernel types. Then, logically, the best kernel is the one that is able to distribute the scores of the images, such that the relevant images have higher scores than the non-relevant ones. Therefore, we suggest the following procedure for selecting the kernel type:

1. after each feedback step use the relevance information given by user to compute for each kernel type K_i the score distribution of the relevant images (ScR_i), and the score distribution of the non-relevant images ($ScNR_i$), here, for $s = 10$ intervals.
2. select as the best kernel $K_i = \min_{K_j} \left(\sum_{s=1}^{10} \min \{ScR_{js}, ScNR_{js}\} \right)$.

Our method of selecting the best kernel tries to fit the score distributions of both relevant and non-relevant images, such that there are as many as possible relevant images with high scores grouped towards the right half of the plot (see Figure 1), and just a few relevant images grouped in the left half side, and vice-versa for the non-relevant images, i.e. there is a good separation between the relevant and non-relevant images. For this, we penalize the relevant images with low scores and the non-relevant images with high scores, and then, we select the best kernel as the one with the smallest penalty (error).

As Zhang and Callan noticed [13], the model is biased, especially for low scoring images, which do not occur between the top (here) 300 images. However, for high scoring images, the model offers a relatively good estimation [13, 9].

However, the above procedure can be applied after each feedback step to select the best kernel type to be used for the following feedback step. Since the procedure is based on the feedback information, it follows that the method might work better when there is more feedback information. Therefore, we suggest to apply the method when using the feedback information accumulated from all previous feedback steps.

As mentioned, previous work [13, 9] used different distributions to model the scores. Then, these distributions were used to predict the behavior of the retrieval model. Instead, in here, we just compute the score distributions based on the information given by user as feedback (relevant and non-relevant images), without using any particular model for these distributions. That is, we do not make any assumption about the behavior of these score distributions.

On the other hand, if we want a way of selecting the best kernel for a given query in advance, we can still use the above procedure. That is, after some feedback we select the best kernel as given by our method, and then, we use the normal and exponential distributions to predict the behavior of the model for the relevant and non-relevant images, respectively. Next, we can calculate the overall (predicted) score distributions for each kernel type by using these models. Finally, we select the kernel that shows the highest (predicted) score distribution of the relevant images and the lowest (predicted) score distribution of the non-relevant images. Note that this procedure uses the score distribution models proposed in the literature for text retrieval [10].

As a conclusion, our method for selecting the kernel type for a particular query is based on score distributions, which are obtained via feedback from user and can be calculated automatically by the system at each feedback step. Next section reports our experimental results.

3 Experiments and Results

3.1 Experimental Setup

For our experiments, we use two test collections of sizes 5000 and 10000, each including 100 relevant images, for each query image. For convenience, we name these sets as 5000_100 and 10000_100. All image collections are quantized to 256 colors in RGB [15] and 166 colors in HSV [14]. We use the same set of 10 images as queries (Q_1,Q_2, ..., Q_{10}) for each experiment.

For evaluation purposes, we use the Test and Control method. The process of obtaining the training and testing sets is described in details in [6]. For each test collection, we create a training set of 300 randomly distributed images. The tests are performed on the the test set. The number of the relevant images within the training and the testing sets for each query is given in Table 1.

Table 1. Number of the relevant images within the training and testing sets.

5000_100	Training	Test
Q_1	9	53
Q_2	6	50
Q_3	4	51
Q_4	8	56
Q_5	7	43
Q_6	5	56
Q_7	4	56
Q_8	4	48
Q_9	5	48
Q_{10}	6	50

10000_100	Training	Test
Q_1	4	62
Q_2	0	49
Q_3	4	52
Q_4	7	46
Q_5	7	47
Q_6	7	46
Q_7	3	44
Q_8	4	46
Q_9	1	49
Q_{10}	2	48

We perform the experiments for a set of 12 kernels: 6 polynomials and 6 radial basis, with general forms given by Equations (3) and (4), respectively. The values of the parameters (a, b, c, and d) and the names of the kernels used in the experiments are presented in Table 2. In all experiments presented in this work, $\sigma = 1$.

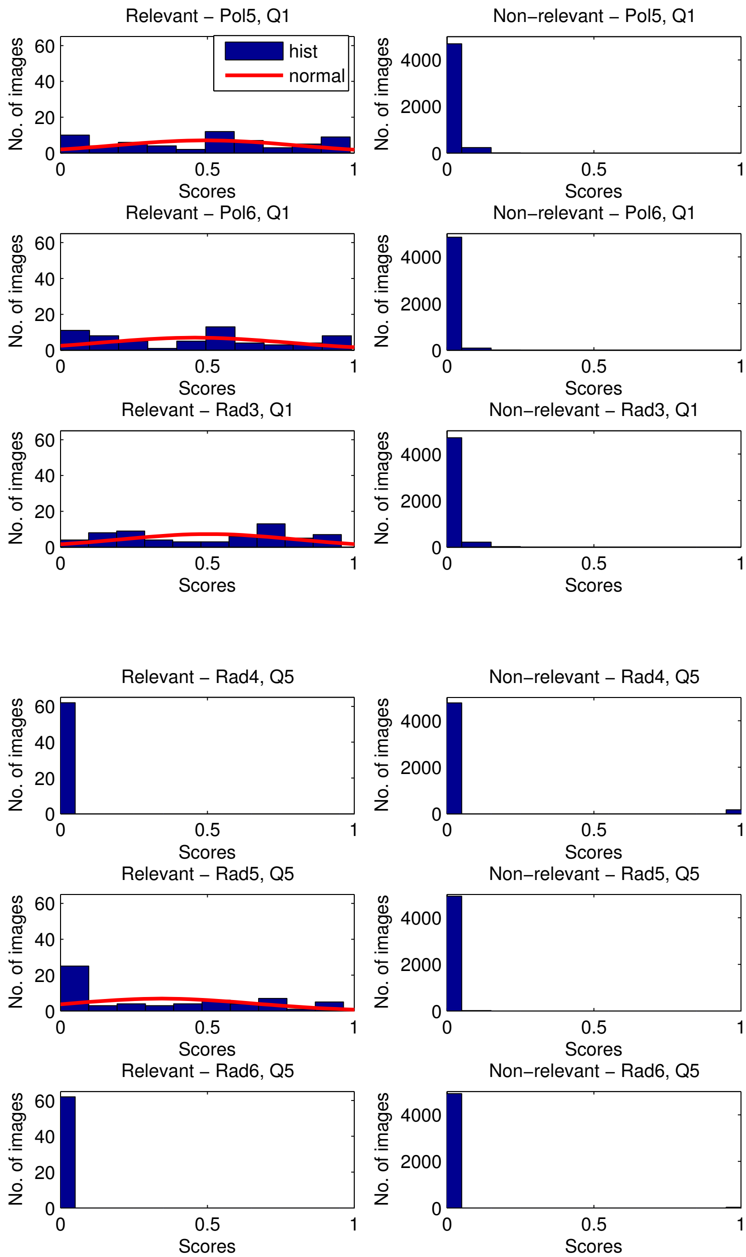

Fig. 1. Score distributions for query Q_1 for 10000_100 in RGB, for $Pol_5, Pol_6, Rad_3, Rad_4, Rad_5, Rad_6$ kernels.

Table 2. Parameters used for the different kernels.

d	Name
$d = 1$	Pol_1
$d = 2$	Pol_2
$d = 3$	Pol_3
$d = 4$	Pol_4
$d = 5$	Pol_5
$d = 6$	Pol_6

a) polynomials

a	b	c	name
$a = 1$	$b = 2$	$c = 1$	Rad_1
$a = 1$	$b = 1$	$c = 1$	Rad_2
$a = 0.5$	$b = 2$	$c = 1$	Rad_3
$a = 0.5$	$b = 1$	$c = 1$	Rad_4
$a = 0.25$	$b = 2$	$c = 1$	Rad_5
$a = 0.25$	$b = 1$	$c = 1$	Rad_6

b) radials

3.2 Discussion of the Results

Figure 1 illustrates the score distributions of the top 300 images for query Q_1 for 6 kernels for 10000_100 collection. The curve shows the normal distribution that fits the score distributions of the relevant images (Figure 1). The other kernels (not shown) present similar score distributions. Also, we obtained similar results for the 5000_100 collection, in both color spaces, for all queries.

As an observation, the score distributions are different for the different kernel types used in the experiments (Figure 1). While all polynomial kernels and Rad_1, Rad_3, Rad_5 kernels fit the score distribution models given in Section (2.1) (i.e. normal distribution for relevant images and exponential distribution for non-relevant images), the other kernels (Rad_2, Rad_4, Rad_6) fit exponential distributions for both relevant and non-relevant images. This is in concordance with the curves these kernels (Rad_2, Rad_4, Rad_6) display in Figure 2, with almost constant R_{norm} values.

Our method of selecting the best kernel tries to fit the score distributions of both relevant and non-relevant images, such that there are as many as possible relevant images with high scores grouped towards the right half of the plot, and less relevant images grouped in the left half side, and vice-versa for the non-relevant images.

For example, in Figure 1, Pol_5 and Pol_6 have very close score distributions, with Pol_6 being slightly better than Pol_5. When comparing Pol_6 with Rad_3, Pol_6 shows more scores of the non-relevant images close to 0 than Rad_3 kernel, but this difference is very small. For the relevant images, Rad_3 performs better, i.e. more relevant images get scores higher than 0. Therefore the winner between the two kernels is Rad_3. A similar discussion can be made to compare Rad_3 and Rad_5, with the same winner Rad_3. Interesting, Rad_4 and Rad_6 kernels show many relevant and non-relevant images grouped together (score close to 0).

By using our procedure given in Section 2.3, the kernels are ordered from the best score distribution to the worst score distribution, as follows: Rad_3, Rad_5, Pol_6, Pol_5, Rad_4, Rad_6. That is, for query Q_1 the best kernel found by our method is Rad_3.

Now, further, we want to check how good our kernel selection method performed. For this, we performed extensive experiments on both collections to see the behavior of the different kernels. To evaluate the quality of the retrieval, we use the R_{norm} measure [3]. For each of the image query, R_{norm} compares the expert provided ranking (known) against the system ranking.

For each query, we plot the R_{norm} values (at each feedback step of 10 images) for each kernel. Space limitation precludes us from showing all the plots obtained from our experiments. However, as an example, in Figure 2 we present the plots of the kernel values obtained for five queries $Q_1, ..., Q_5$ for 10000_100 image test collection in RGB color space.

From Figure 2, one can notice that Rad_5 and Rad_3 kernels show very similar curves, which display significantly higher results than the curves corresponding to the other kernel types. That is, these two kernels are a better choice than the other kernels. However, if one considers the efficiency issue, then the Rad_3 kernel should be preferred over Rad_5, since it might reduce the computational cost during retrieval. That is, from our experiments, the retrieval system should use the Rad_3 kernel to achieve best effectiveness and efficiency.

This result is consistent with the result obtained by using our score distribution procedure for selecting the best kernel (same selected kernel type). That is, our method could be a viable solution to automatically select the kernel type in an AIRS.

4 Conclusions and Future Work

Kernel methods offer an elegant solution to increase the computational power of the linear learning algorithms by facilitating learning indirectly in high-dimensional feature spaces. Therefore, kernels are important components that can improve the retrieval system.

Previous work [9, 10, 8] modeled the score distributions of search engines for relevant and non-relevant documents on a per query basis. Our previous work [6, 7] revealed that there is no overall best kernel, but just maybe a best kernel for each query. Based on these works, we propose a kernel selection procedure that automatically chooses the best kernel type for a query by using the score distributions of the relevant and non-relevant images given by user as feedback.

For testing our method, several kernels having polynomial and Gaussian Radial Basis Function (RBF) like forms (6 polynomials and 6 RBFs) are applied to generic images represented by color histograms in RGB and HSV color spaces. We test our method on two image collections of sizes 5000 and 10000 and select the best kernel corresponding to each query. Then, from extensive experimental results on these collections we select the desired kernel to be used for each query. Our method found the same best kernel type for a query as the desired kernel displayed by our extensive experiments.

As future work, we plan to investigate the usage of our kernel selection method with other kernel-based learning methods such as Kernel Perceptron and Kernel SVMs. Also, one can study the applicability of the different mixture models of the score distributions to the selection of the best kernel for a particular query in an image retrieval system.

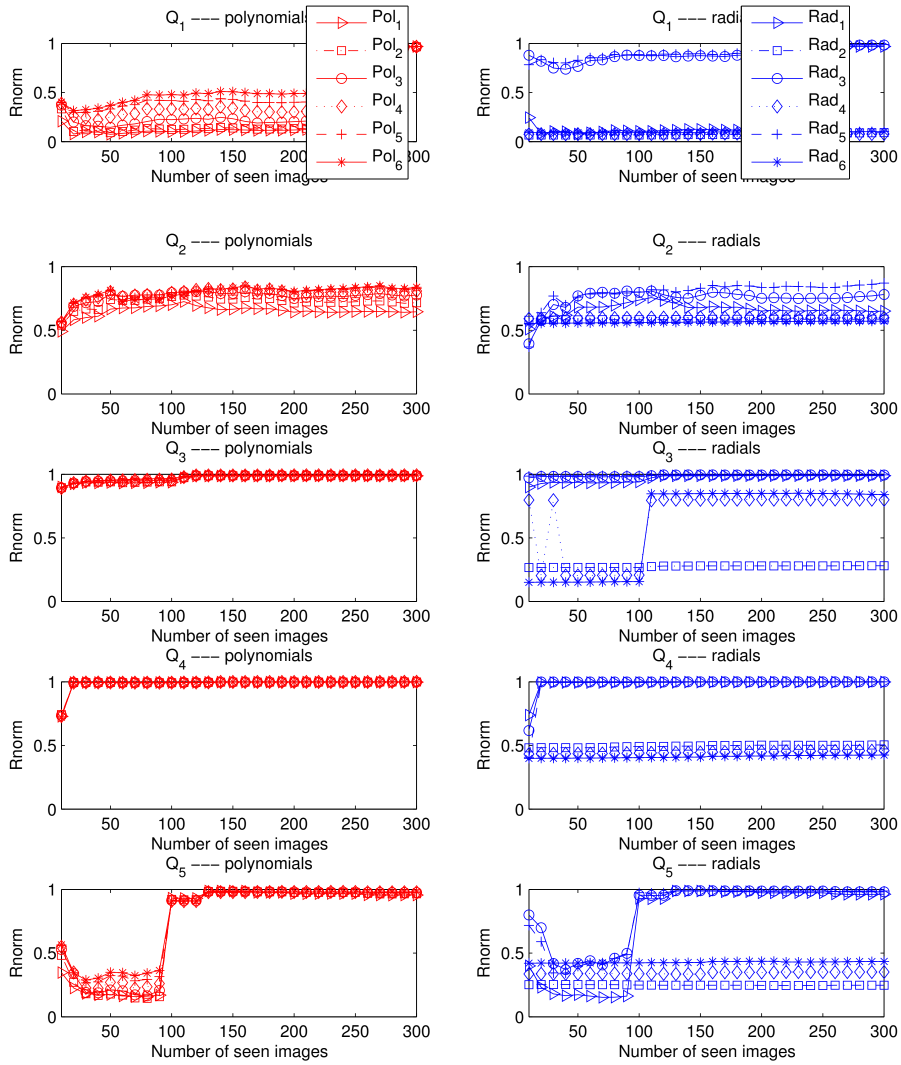

Fig. 2. Kernel results for query Q_1, Q_2, Q_3, Q_4, Q_5 for 10000_100 in RGB.

References

1. Doloc-Mihu, A., Raghavan, V., V., Bollmann-Sdorra, P.: Color Retrieval in Vector Space Model. Proceedings of the 26th International ACM SIGIR Workshop on Mathematical/Formal Methods in Information Retrieval, MF/IR (2003)
2. Raghavan, V., V., Wong, S. K. M.: A Critical Analysis of Vector Space Model for Information Retrieval. Journal of the American Society for Information Science **37** (1986) 279-287
3. Bollmann-Sdorra, P., Jochum, F., Reiner, U., Weissmann, V., Zuse, H.: The LIVE Project - Retrieval Experiments Based on Evaluation Viewpoints. Proceedings of the 8th International ACM SIGIR Conference on Research and Development in Information Retrieval (1985) 213–214
4. Chapelle, O., Haffner, P., Vapnik, V.: SVMs for Histogram-Based Image Classification. IEEE Transactions on Neural Networks **5** (1999) 1055–1064
5. Huijsmans, D. P, Sebe, N.: How to Complete Performance Graphs in Content-Based Image Retrieval: Add Generality and Normalize Scope. IEEE Transactions on Pattern Analysis and Machine Intelligence **27** (2005) 245–251
6. Doloc-Mihu, A., Raghavan, V., V.: Selecting the Kernel Type for a Web-based Adaptive Image Retrieval System (AIRS). Internet Imaging VII, Proceedings of SPIE-IS&T Electronic Imaging, SPIE **6061** (2006) 60610H
7. Doloc-Mihu, A., Raghavan, V., V.: Using Score Distribution Models to Select the Kernel Type for a Web-based Adaptive Image Retrieval System (AIRS). International Conference on Image and Video Retrieval (CIVR 2006), LNCS, Springer, July 13-15, 2006 (to appear).
8. Swets, J., A.: Information Retrieval Systems. Science **141** (1963) 245-250
9. Manmatha, R., Feng, F., Rath, T.: Using Models of Score Distributions in Information Retrieval. Proceedings of the 24th ACM SIGIR Conference on Research and Development in Information Retrieval (2001) 267-275
10. Arampatzis, A., Beney, J., Koster, C., van der Weide, T., P.: Incrementally, Half-Life, and Threshold Optimization for Adaptive Document Filtering. The 9th Text REtrieval Conference (TREC-9) (2000)
11. Cristianini, N., Shawe-Taylor, J.: An introduction to Support Vector Machines and other kernel-based learning methods. Cambridge University Press (2000)
12. Tipping, M., E.: The Relevance Vector Machine. Advances in Neural Information Processing Systems **12** (2000) 652-658
13. Zhang, Y., Callan, J.: Maximum Likelihood Estimation for Filtering Thresholds. Proceedings of the 24th ACM SIGIR Conference on Research and Development in Information Retrieval (2001)
14. Smith, J., R.: Integrated Spatial and Feature Image Systems: Retrieval, Analysis and Compression. PhD thesis, Columbia University, 1997
15. Heckbert, P., S.: Color Image Quantization for Frame Buffer Display. Proceedings of SIGGRAPH **16** (1982) 297-307