Example 1.

From Table 2 it is clear that the use of three classes instead of two gives a higher average precision. Given a similar number of user rated songs for training, the three-class method outperforms the two-class method. This can for example be observed when comparing $N = 5$ in the 3-class method (using $5 \times 3 = 15$ songs for training) and $N = 7$ in the 2-class method (using $2 \times 7 = 14$ songs for training) using cosine similarity. In this case, the retrieval performance of the 3-class method is 20% better using a training set that is 7% larger.

Table 2: Mean average precision of retrieved data in preliminary experiments. The average of all average precision values were calculated for each training data set size N. The Random rows show the average precision for a randomly sorted retrieval list.

Num. of classes	Similarity	Amount of training data (N) 1	3	5	7	9
2	Random	0.241	0.223	0.229	0.216	0.197
	Cosine	0.282	0.337	0.352	0.331	0.332
	Entropy	0.288	0.332	0.357	0.361	0.349
3	Random	0.236	0.234	0.246	0.240	0.234
	Cosine	0.305	0.364	0.370	0.388	0.389
	Entropy	0.268	0.288	0.351	0.424	0.392

The results in Table 2 show that all retrieval experiments achieved higher mean average precision than the random results, proving the effectiveness of the TreeQ method. Comparison of the results of the similarity measures *cosine* and *entropy* indicates that the cosine similarity is slightly better than the symmetric relative entropy in the 3-class experiments.

4.3.3 Ambiguity in retrieval performance

Example 2.

The results in Table 4 show the mean average precision for each test subject using both scoring method *Sim* and *DiffSim*. Comparison of the results do not indicate any significant difference between scoring methods, although *DiffSim* has achieved higher performance for subject U_2. Moreover, no clear correlation is observed between the amount of training data and average precision.

In order to show how the variations in performance depend on how the training data set is selected, the results within each setting of N have also been presented. Figures 7 and 8 illustrate the maximum, minimum, and mean average precision for each set of experiments per user, based on scoring method *Sim* using cosine similarity.

Say why you choose to show results in a certain way

Table 4: Mean average precision of different ranking methods

Subject	Amount of training data (N) 1	3	5	7	9
$U_1(Sim)$	0.335	0.426	0.402	0.463	0.469
$U_1(DiffSim)$	0.273	0.400	0.378	0.506	0.378
$U_2(Sim)$	0.276	0.302	0.338	0.314	0.310
$U_2(DiffSim)$	0.296	0.370	0.411	0.410	0.376

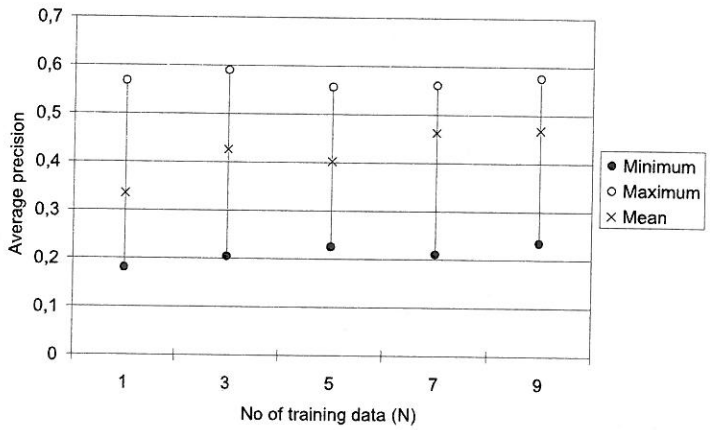


Figure 7: Average precision of all experiments (Subject U_1)

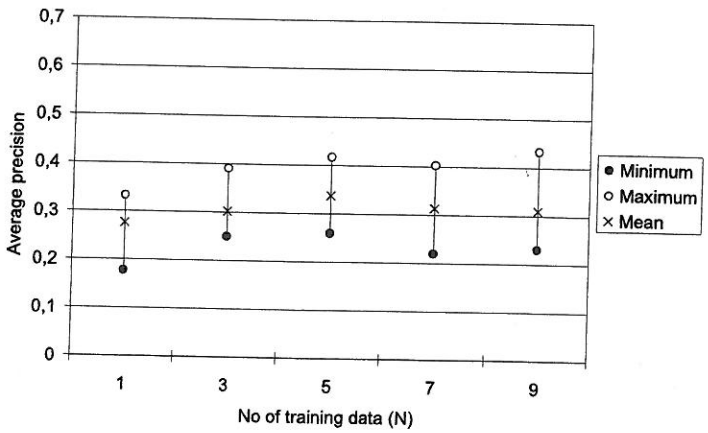
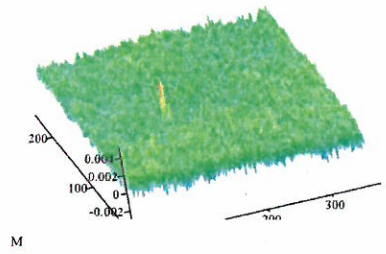
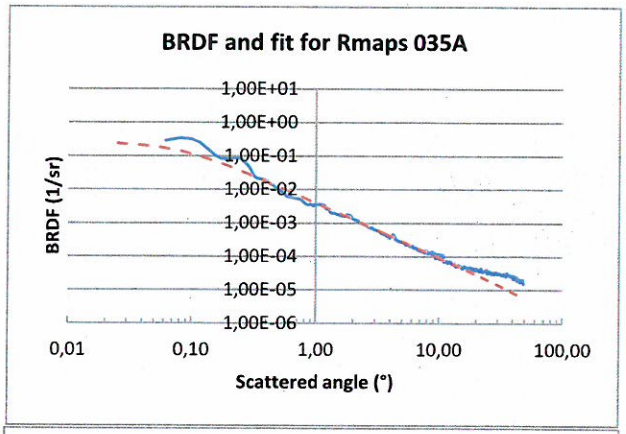


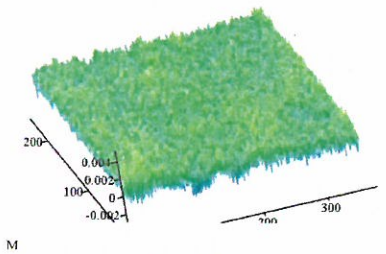
Figure 8: Average precision of all experiments (Subject U_2)

As clear from Figures 7 and 8, there is a wider range of variations in average precision for subject U_1 compared to subject U_2. The minimum average precision is similar for both subjects, but the maximum average precision is significantly higher for subject U_1. Furthermore, no clear correlation is observed between the amount of training data and mean average precision, as the results in Table 4 also have shown.

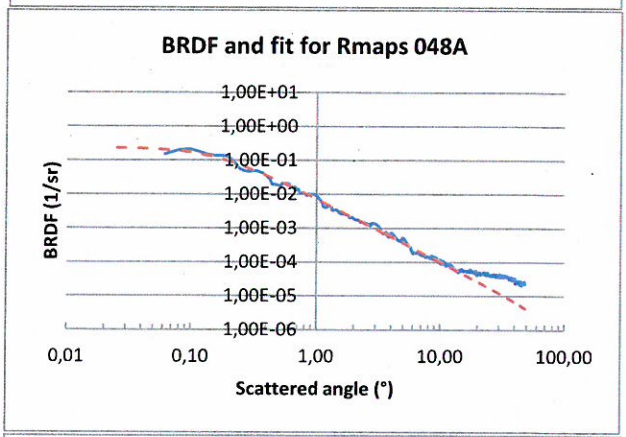


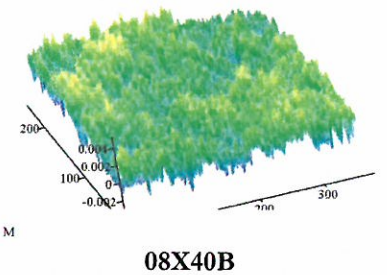
035X40A



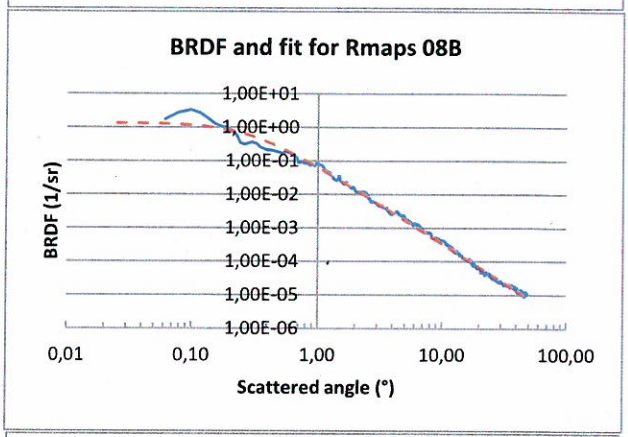


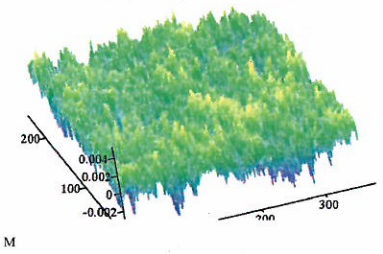
048X40A



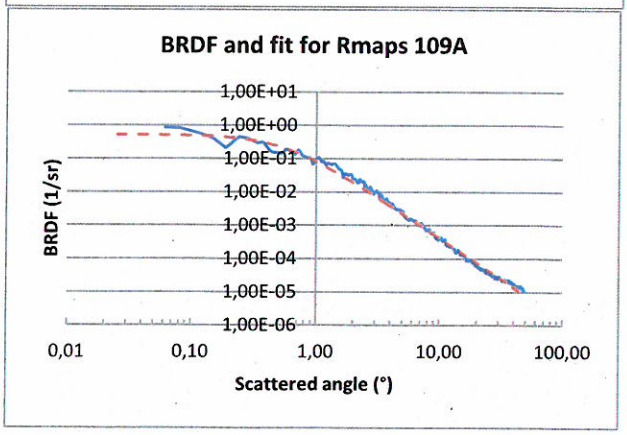


08X40B





109X40A



Sample	Surface	BRDF
035	035A	035X10A
		035X40A
	035B	035X10B
		035X40B
048	048A	048X10A
		048X40A
	048B	048X10B
		048X40B
08	08A	08X10A
		08X40A
	08B	08X10A
		08X40A
097	097A	097X10A
		097X40A
	097B	097X10B
		097X40B
109	109A	109X10A
		109X40A
	109B	109X10B
		109X40B

Figure 3-26. 3D-views of the set of selected SiC CVD samples, their representative BRDFs, and their denominations

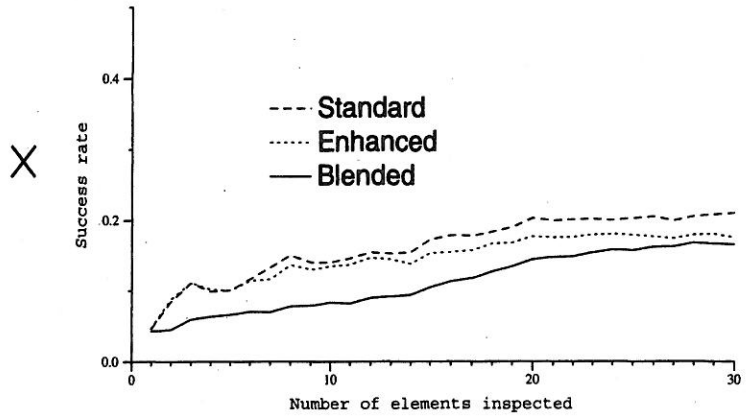


FIGURE 7. *Success rate as the number of inspected items is increased. It is clear that blending is not effective.*

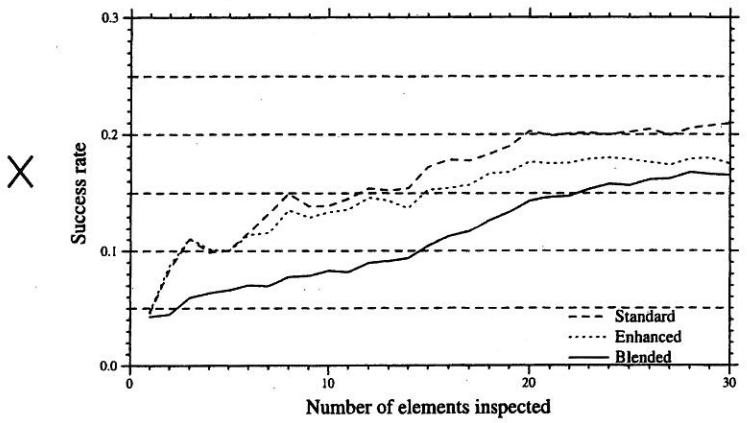


FIGURE 7. *Success rate as the number of inspected items is increased. It is clear that blending is not effective.*

Badly designed graphs. These graphs show the same data. In the upper version, poor use has been made of the vertical space available, and the legend is awkwardly placed. Fonts and size are changed unnecessarily, and are inconsistent with the main text. In the lower version, the vertical scaling and fonts have been corrected, but unnecessary ornamentation has been introduced. The grid lines and heavy border now greatly outweigh the data being presented.

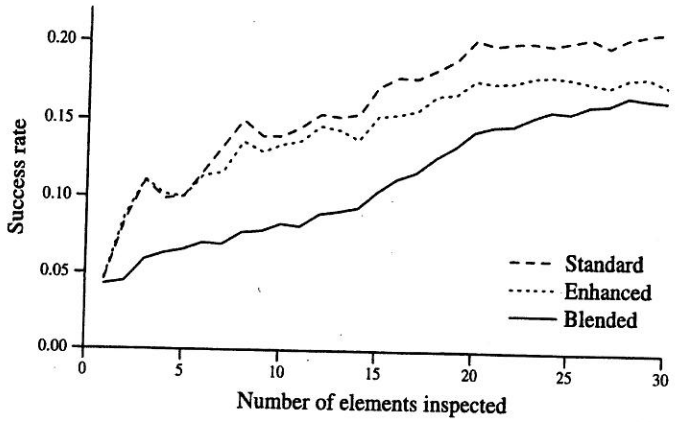


FIGURE 7. *Success rate as the number of inspected items is increased. It is clear that blending is not effective.*

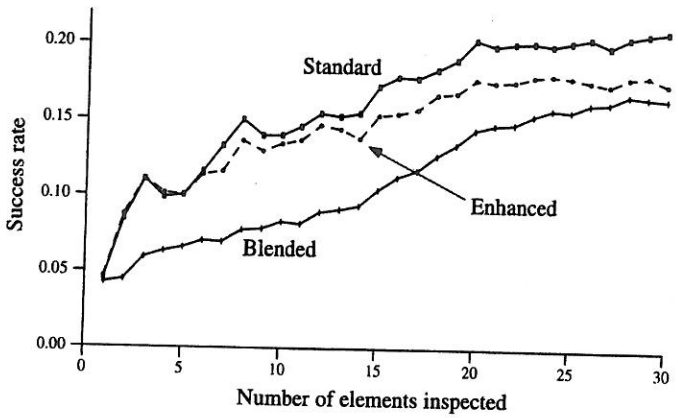


FIGURE 7. *Success rate as the number of inspected items is increased. It is clear that blending is not effective.*

Graphs reconsidered. These graphs show the same data as those on the previous page. Vertical scale is now completely corrected, and unnecessary tick marks have been removed. In the lower version, the data lines are stronger and the legend has been replaced with direct labelling. Line ticks have been introduced to reflect the fact that the data is discrete, that is, non-integer values are not meaningful.

STATISTICS	SMALL	LARGE
Characters	18,621	1,231,109
Words	2,060	173,145
After stopping	1,200	98,234
Index size	1.31 Kb	109.0 Kb

TABLE 6. *Statistics of text collections used in experiments.*

	Collection	
	Small	Large
File size (Kb)	18.2	1,202.3
Index size (Kb)	1.3	109.0
Number of words	2,060	173,145
After stopping	1,200	98,234

TABLE 6. *Statistics of text collections used in experiments.*

Two versions of a table. The upper version is poor. No use has been made of table hierarchy—all the elements are at the same level, so that case has to be used to differentiate between headings and content. Different units have been used for file sizes in different lines (assuming characters are one byte each). Units haven't been factored out in the last line and the precision is inconsistent. The heading of the first column is unnecessary and the table has too many horizontal lines.

In the lower version there are no vertical lines. Rows of the same type are now adjacent so that they can be compared by the reader. Note that the values of different units do not need to be vertically aligned on the decimal point or presented with the same precision.