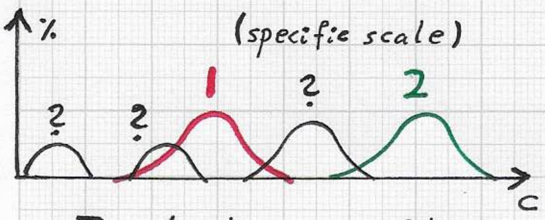
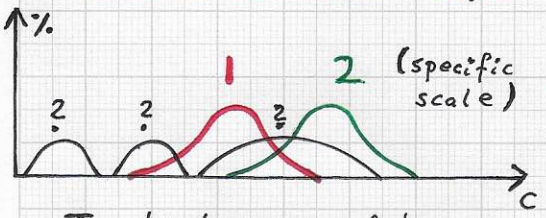
■ OBJECT AND MATERIAL EIGENFUNCTIONS – Cont'd.

• Laplacian eigenfunctions and neural networks: ...

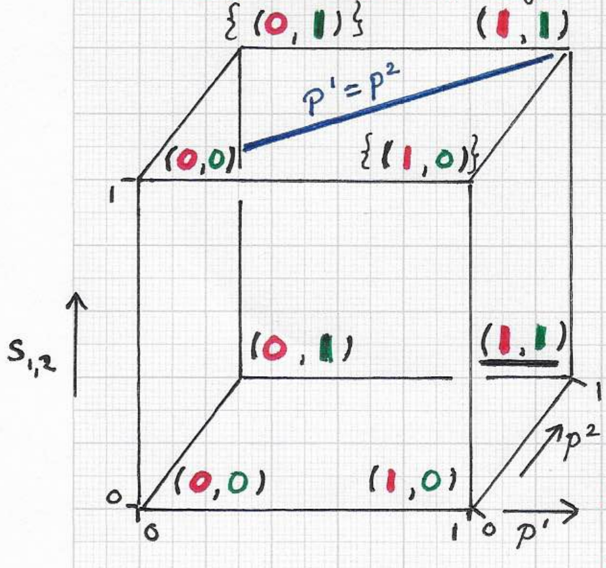
Overlap possibilities for classified histograms and one unclassified histogram:



Two histograms of known classes, not overlapping:



Two histograms of known classes (1, 2), overlapping.



Visualization of possible decider function tuples (F^1, F^2). The numerical values (1,0) and (0,1) are not possible for $s_{1,2}=1$, indicated by '{...}'.

The two top figures (left) show one scenario where the histograms of two known classes do not overlap and therefore have a degree-of-similarity value of **0, and** another scenario where the histograms of the two known classes **1** and **2** do overlap and have a degree-of-similarity value of $s_{1,2} > 0$. The histogram of the unclassified data is the '?' histogram, "moving from left to right." In the first scenario, the (p^1, p^2) tuple has values $(0, 0)$, $(p^1 > 0, 0)$ and $(p^1 > 0, p^2 > 0)$. The same sequence of value tuples results in the second scenario. One can assume that $p^1, p^2 \in [0, 1]$. For example, one could consider to use simple, linearly varying decider functions as default functions, i.e., $F^1(p^1, s_{1,2}) = p^1$ and $F^2(p^2, s_{1,2}) = p^2$. **Even when $s_{1,2} = 0$ it is imaginable that $p^1 = p^2 = 1$ (subject to using a degree-of-similarity measure producing this result). For $s_{1,2} = 1$, it is logically necessary that $p_1 = p_2$** (blue line on top face).

...

OBJECT AND MATERIAL EIGENFUNCTIONS – Cont'd.

- Laplacian eigenfunctions and neural networks:...

Singular, extremal case of degree-of-similarity of histograms:

Class-histograms 1 and 2 are identical, and they are also identical to '?'.

At this scale, one cannot distinguish between material classes 1 and 2. Thus, when an unclassified image segment has a degree-of-similarity of a certain value with one of the two classes it must have the same degree-of-similarity value with the other class.

- **Note.** Until now, this discussion has not explicitly covered the fact that a multitude of material samples is involved in generating a large sample database in a training phase. Each material sample is represented by **H** scale-specific coefficient histograms, and each material class is captured by many samples (image segments) to cover the variation of coefficient value histograms for each class.

- **Note. The** logical explanation for the p-value requirements is the following consideration:

$$(s_{1,2} = 1 \wedge p^1 = \ldots) \Rightarrow p^2 := p^1,$$
$$(s_{1,2} = 1 \wedge p^2 = \ldots) \Rightarrow p^1 := p^2.$$

In other words, p-values MUST be identical when the two histograms of the two known classes are identical (for a specific scale). The left figure (top) illustrates an extremal case: the three coefficient histograms for the known classes 1 and 2 and '?' are identical. Thus, $p^1 = 1$ and $p^2 = 1$, **implying that '?' is of class 1 and of class 2, with a probability of 1 for each class.** (This implication concerns the specific scale used for comparison. Nevertheless, in principle, this extremal case could simultaneously arise for all scales.)

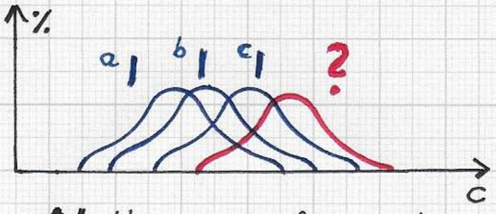
One can handle such a "contradiction" by defining the default decider functions as $F^1 = p^1/2$ and $F^2 = p^2/2$ (with 2 being the number of known – PHYSICALLY DIFFERENT – classes).

...

■ OBJECT AND MATERIAL EIGENFUNCTIONS – Cont'd.

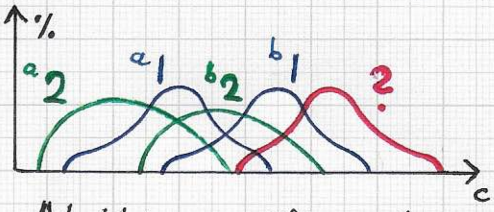
• Laplacian eigenfunctions and neural networks:...

Class represented by several class-samples with associated sample-specific histograms:



At this specific scale, material class 1 is represented by three coefficient value histograms, a, b and c, associated with three same-class samples. One must define and calculate p-values and an F-value for the unclassified '?' histogram.

Two-class example: c-domains of histograms overlap, but distributions of c-values differ significantly:



At this specific scale, the histograms for the two material classes 1 and 2 (for two samples a and b) have significantly different "shapes." The shape of '?' is qualitatively much more similar to the shape of $^a1$ and $^b1$. **Statistically it can be expected that p-values of '?' are higher for class 1.**

One important aspect has been left out in this discussion: Each material is characterized by MANY training samples in the "database." The left figure shows an example where material class 1 is represented by three samples, i.e., coefficient histograms $^a1$, $^b1$ and $^c1$ (for a specific scale). An unclassified histogram '?' must be classified; '?' overlaps with all three sample histograms belonging to class 1. Thus, the three p-values calculated for '?' (when compared with $^a1$, $^b1$ and $^c1$) are all greater than 0. In this situation, it is reasonable to use the largest of the three degree-of-similarity p-values as "final" p-value for '?' and assign an F(p)-value to '?'. The figure (left, bottom) high-lights the case of four histograms (two classes, 1 and 2, two samples each, a and b) that overlap in the c-domain. Despite this overlap, '?' should only be capable of having high p-values with class-1 histograms.

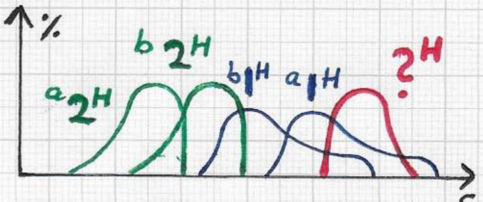
...

■ OBJECT AND MATERIAL EIGENFUNCTIONS — Cont'd.

• Laplacian eigenfunctions and neural networks: ...

Classification steps performed for an unclassified image segment should be performed in a "prioritized" fashion – to obtain a classification result in "minimal time":

Scale-1 comparison: The unclassified histogram "?'" can be compared with two classes (1, 2), each represented by two samples (a, b).



Scale-H comparison: The unclassified histogram '?H' can be compared with two classes (1, 2), each represented by two samples (a, b).

• **Priority/optimization:**

i) classes ordered with respect to frequency of occurrence – e.g., class 1 most common

ii) for each class, scales are ordered based on their discriminative power

⇒ Perform comparisons accordingly, following this order!

• **Note.** Another algorithm design aspect is crucial for ensuring best-possible efficiency when comparing histograms and computing p-values, for example, for an unclassified image segment:

(i) Historically, certain material classes are found more often than other classes. Consequently, one should make sure that the most common classes are given highest priority for comparison. Thus, classes should be ordered/indexed based on their frequency.

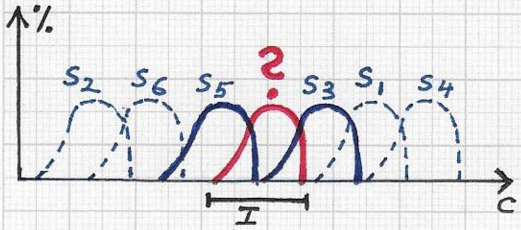
(ii) Considering a specific material class, one can assume that certain scales have more discriminative power than other scales. Thus, one should also design an algorithm and data structures in such a way that scales with maximal "recognition power" are used first in the histogram comparison process. Scales should be ordered accordingly.

...

■ OBJECT AND MATERIAL EIGENFUNCTIONS - Cont'd.

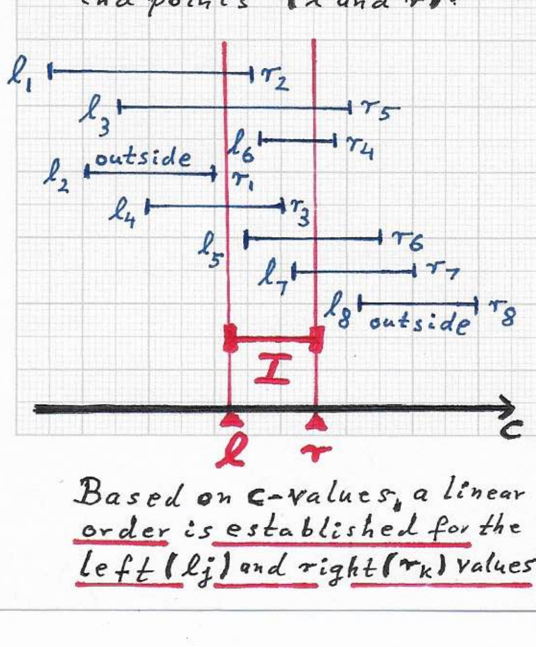
• Laplacian eigenfunctions and neural networks:...

Domain intervals of all histograms must be ordered for rapid identification of class histograms to be used for degree-of-similarity calculations:



For this specific class and scale six sample histograms are provided, $s_1, \dots, s_6$. The unclassified histogram '?' has the interval I as its domain on the c-axis. Only s_3 and s_5 have domains that intersect with I. Thus, only s_3 and s_5 need to be used for comparison.

Sorting c-domain intervals I, considering left and right end points (l and r):



Based on c-values, a linear order is established for the left (l_j) and right (r_k) values.

In addition to determining an order for the material classes and for their respective scale-specific histograms, one must also optimize the search for the specific sample histograms (per class and per scale) that are necessary for comparisons. For example, if 1000 sample histograms are available (for a certain scale of a specific class), then one will have to determine quickly the subset of these 1000 histograms one must use to calculate degree-of-similarity values for a same-scale histogram of the unclassified image segment. The left figure illustrates the search problem. For the interval of interest (I) with end points/values l and r, one must quickly determine the sample-specific intervals of s_i that are entirely to the left or right of I. The sorted left (right) interval values are $l_1, \dots, l_8$ ($r_1, \dots, r_8$). In the shown scenario, one sees that $r_1 < l$ and $l_8 > r$; thus, $[l_2, r_1]$ and $[l_8, r_8]$ are "outside."

...