■ OBJECT AND MATERIAL EIGENFUNCTIONS – Cont'd.

• Laplacian eigenfunctions and neural networks:...

Minimizing search time via linearly sorted histograms/histogram domain intervals and optimal binary search:

number of histograms (N)	$\log_2(N)$
128	7
512	9
1,024	10
4,096	12
8,192	13

Coefficient value histograms have non-zero %-values only for specific coefficirnt value (c-value) intervals. IT IS ASSUMED HERE THAT THESE INTERVALS ARE **NOT** MAPPED TO THE UNIT INTERVAL [0,1] for data analysis and classification purposes.

The table shows that substantial efficiency gain is possible by employing binary search – allowing one to identify histograms needed for data comparison in logarithmic time.

For example, if the number of histograms used for comparison is 1,024, on average, binary search is 100 times more efficient than linear search.

Whenever the number of sample histograms (for a specific class and scale) is "large" one should sort the c-domain intervals of all samples using l-values and r-values separately; thus, a first ordering pass sorts all l-values and a second ordering pass sorts all r-values. A sample histogram's c-daimain has a non-empty intersection with the interval $I = [l, r]$ only if $l_j \leq r \wedge r_k \geq l$, where the sample histogram's c-domain interval is $[l_j, r_k]$. (The indices j and k reflect the from-left-to-right ordered l- and r-values.) **Thus, one must search for the minimal index j_{min} for which $r < l_{j_{min}}$ and maximal index k_{max} for which $l > r_{k_{max}}$ to determine all "outside-I" sample histogram intervals**, see figure on previous page (left, bottom). **Sorting is a pre-processing step; searching is a "real-time" step, done during classification.**

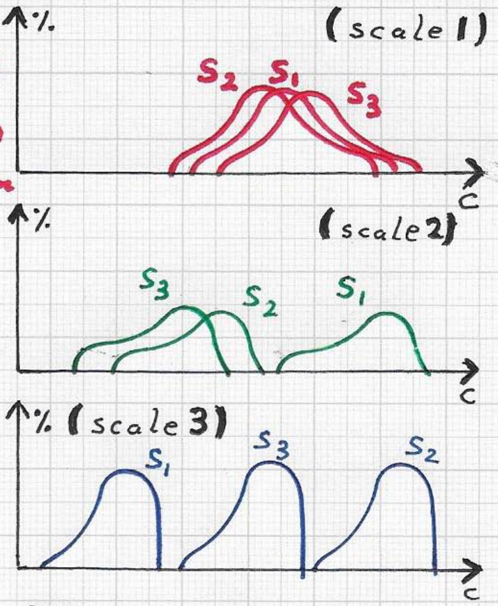
...

■ OBJECT AND MATERIAL EIGENFUNCTIONS – Cont'd.

• Laplacian eigenfunctions and neural networks:...

Scale-specific coefficient value histograms of class samples are more or less "indicative" for a material class:

nearly 'one unique' histogram



Three coefficient value histograms for three scales associated with three samples s_1, s_2 and s_3 of one specific class.

⇒ **Rule (here):**

"This class has 'one unique' histogram for scale 1.

If a new, unclassified histogram for scale 1, of a new given image segment, differs from the 'one unique' histogram, then it will NOT belong to that class."

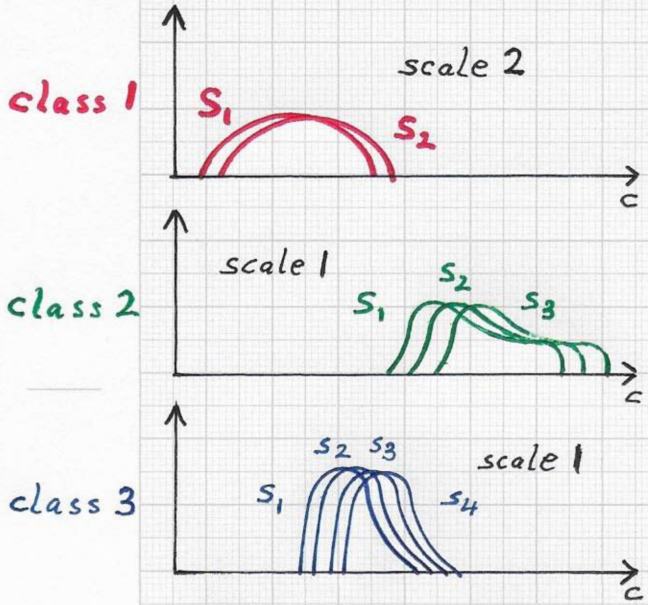
• **Note.** The figures illustrate another aspect one should consider and take advantage of, for the purpose of most efficient classification. The three images (left) provide sketches of possible histogram behavior for three segments (s_1, s_2, s_3) for three scales. The three histograms (per scale) are "maximally concentrated" for scale 1 and "maximally spread out" for scale 3 (for this specific class represented by these specific samples). **One can define an overall measure D to assess how "concentrated" the sample histograms are for a particular scale. For example, the sum $D = d_{1,2} + d_{1,3} + d_{2,3}$, where $d_{i,j}$ measures histogram pair difference, is minimal for scale 1 for the sketched scenario.** One can and should utilize this insight. The fact that one scale exhibits this "maximal concentration" of histograms for a certain class makes this scale "class-typical."

...

OBJECT AND MATERIAL EIGENFUNCTIONS — Cont'd.

• Laplacian eigenfunctions and neural networks:...

Three classes, represented by different numbers of samples, and three different scale histograms:



Three classes — 1, 2, 3 — where the class index defines the order of frequency-of-occurrence, with class 1 being most frequent. The three classes are represented in the database by 2, 3 and 4 samples, respectively. Based on "measure D" — concentration of coefficient value histograms — scales 2, 1 and 1 are sketched here, associated with the respective classes; these sketched scale-specific data are assumed to be the sets of most concentrated histograms, minimizing the D-value for these three classes.

In summary, we can use "class priority" and "scale priority" to design a highly efficient data analysis and classification pipeline and system modules that can generate a classification result effectively. For example, material classes of interest (threat classes) are more or less common; thus, class 1 should be the most common class etc. The classes stored in the database of all classified sample data can be linearly ordered accordingly — and this order should imply the order of comparing the classified sample data with a given unclassified image segment. In addition, each class has a class-specific scale that minimizes the "measure D" (last page) — the measure that is the sum of all pairwise differences between coefficient value histograms. Thus, for a specific class, one should (first) consider the scale with minimal D-value for classification.

...

■ OBJECT AND MATERIAL EIGENFUNCTIONS – Cont'd.

• Laplacian eigenfunctions and neural networks: ...

• **Classification strategy involving ELIMINATION of classes:**

• For all class and scale sample data in the database do the following:

→ Identify those classes that have sets of **coefficient value histograms** – for one or several scales – that are exhibiting a high degree of concentration in the sense of the **D-measure**.

→ Represent and store these histogram concentration data in the database as well.

• For a new, unclassified image segment, ELIMINATE classes as potentially **matching** classes:

→ Compare all classes in the database with "high-concentration" histograms (for certain scales) with the relevant scale (or scales) of the unclassified segment.

→ **Eliminate all classes that cannot be class matches based on the comparison with sets of high-concentration histograms.**

We must recall that these "most concentrated histograms" allow one to determine rapidly whether an unclassified image segment does NOT belong to a specific class. Considering the illustrated three-class scenario from the previous page, one might obtain the following result: the scale-2 histogram of an unclassified image segment has **LARGE** difference values for all class-1-scale-2 histograms, and, also, the scale-1 histogram of this unclassified image segment has **LARGE** difference values for all class-2-scale-1 and all class-3-scale-1 histograms. **Thus**, the unclassified segment is **NOT** a match for any of the three classes of interest. Generally, one can assume that several classes stored in the sample database have scales for which histograms are highly concentrated. The strategy described here (left) can be used to reduce the number of possibly matching classes.

...

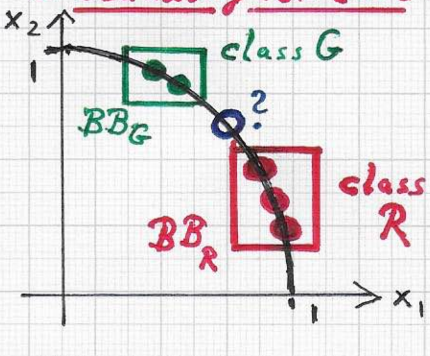
■ OBJECT AND MATERIAL EIGENFUNCTIONS – Cont'd.

• Laplacian eigenfunctions and neural networks: ...

• **Class-0** image segments are ALL segments that do NOT represent any of the threat material classes of interest.

Class-0 segments should be recognized with minimal computational effort.

• Geometrically, one can understand a discrete, i.e., binned, normalized histogram as a point in a space that has the number of bins as its dimension. **Thus, one can employ a bounding box test:**



- All points represent histograms for a specific scale, where $x_i \geq 0 \wedge \sum x_i^2 = 1$.
- Class R is represented by **3** sample segments (⇒ **3** histograms, **3** points).
- Class G is represented by **2** sample segments (⇒ **2** histograms, **2** points).
- Point O (?) is unclassified and is outside the cuboids/bounding boxes BB_R and BB_G. ⇒ Point O is a class-0 histogram point. ⇒ **Irrelevant** segment.

Computing and storing the values of the **D-measure** provides a means for eliminating material classes quickly from a more complex and "time-consuming" analysis and classification procedure of a given unclassified image segment. **The D-measure** makes it possible to rapidly answer the question **"Is the unclassified image segment NOT of this class?"** – answering this question for all classes that have (at least) one scale for which their sets of coefficient value histograms are highly concentrated. **Only material classes for which the answer to this question is 'NO' will be considered for detailed analysis and classification steps** – potentially resulting in one or multiple high-probability class-matches for the unclassified segment. **CLASS 0** – the class that **DOES NOT** represent a threat, a class of interest – **should be recognized immediately** for efficiency.

...