Stratovan

DISTRIBUTION FEATURES AND CLASSIFICATION

- Guiding facts, principles and thoughts:

1) SMALL DATA sets have advantages — storing, transferring, processing, analyzing etc. can be done efficiently.
2) Our "objects" are sets of voxels defining a specific material. Moreover, each material class is (typically) defined by MULTIPLE OBJECT SAMPLES — with their individual features / feature value distributions.
THUS, FOR EACH MATERIAL CLASS ONE HAS MULTIPLE SAMPLES (individual segments), WITH EACH SAMPLE DEFINING ITS OWN DISTRIBUTIONS OF KNOWN OR COMPUTED (derived) FEATURE VALUES. SINCE EACH OF THESE DISTRIBUTIONS HAS ITS OWN CHARACTERISTICS (e.g., μ, σ^2), ONE CAN CONSIDER THE "DISTRIBUTION OF CHARACTERISTICS OF ALL SAMPLE-SPECIFIC DISTRIBUTIONS."
3) One establishes a MODEL of a material class. This model can be defined via LEARNING the material's feature values by considering MULTIPLE SAMPLES. Each sample (object/segment)

...

■ DISTRIBUTION FEATURES – Cont'd.

3) ... produces several histograms/distributions of feature values. Considering only the mean μ and variance σ^2, each feature and each sample define a "feature-and-sample-specific histogram/distribution" with ITS OWN (μ, σ^2) CHARACTERISTIC. By considering ALL SAMPLES USED FOR LEARNING OF A SINGLE MATERIAL CLASS, ONE OBTAINS DISTRIBUTIONS OF (μ, σ^2) TUPLES, all generated empirically.

4) In addition to analyzing the distributions of (empirical) (μ, σ^2) tuples, one could compute and analyze the distributions of the 1st, 2nd, 3rd and 4th MOMENTS of multiple sample-specific distributions.
(BUT: The goal is to use the smallest number of "fingerprint data" necessary to separate multiple distributions from each other.)

5) Simple example – data and complexity:
- No. of materials: 10
- No. of samples/material: 100
- Feature considered: ρ (=density)
- Distribution characteristic: (μ, σ^2)

$\Rightarrow$ Each material has an associated distribution of 100 (μ, σ^2)-tuples. $\Rightarrow$ ...

$\Rightarrow$ 1000 ρ-value distributions, each characterized by its (μ, σ^2) tuple

■ DISTRIBUTION FEATURES - Cont'd.

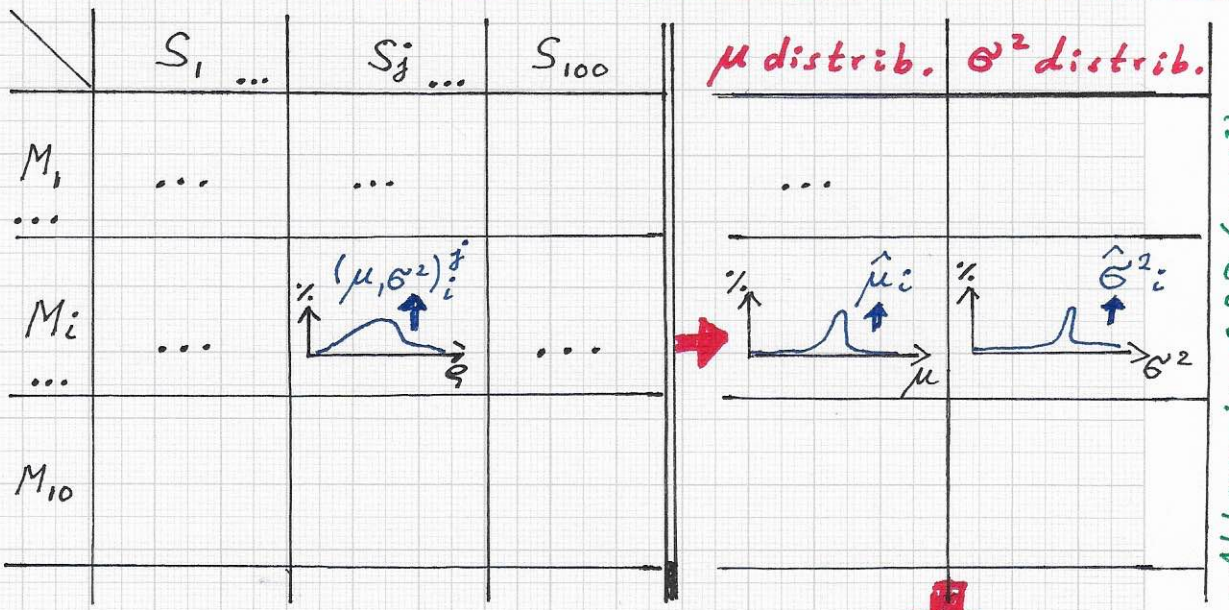
5) Example...

⇒ Each material-specific distribution of (values of) (μ, σ^2)-tuples has its own $(\hat{\mu}, \hat{\sigma}^2)$ "characteristic."

⇒ TEN $(\hat{\mu}, \hat{\sigma}^2)$-TUPLES RESULT FOR TEN MATERIALS. ⇒ **SMALL DATA.**

(The numbers of features used to characterize material properties and to characterize distribution properties [moments] must be large enough to classify materials "properly") subject to classification quality requirements - BUT THESE NUMBERS SHOULD NOT BE LARGER!)

• Illustration: "Distributions of distribution characteristics"



• Alternative: 2D (μ, σ^2) distributions

➡ materials: $M_1, \ldots, M_{10}$

➡ samples: $S_1, \ldots, S_{100}$

$\{(\hat{\mu}, \hat{\sigma}^2)_i\}_{i=1}^{10}$ **ONLY 10 TUPLES**