Bernd Hamann

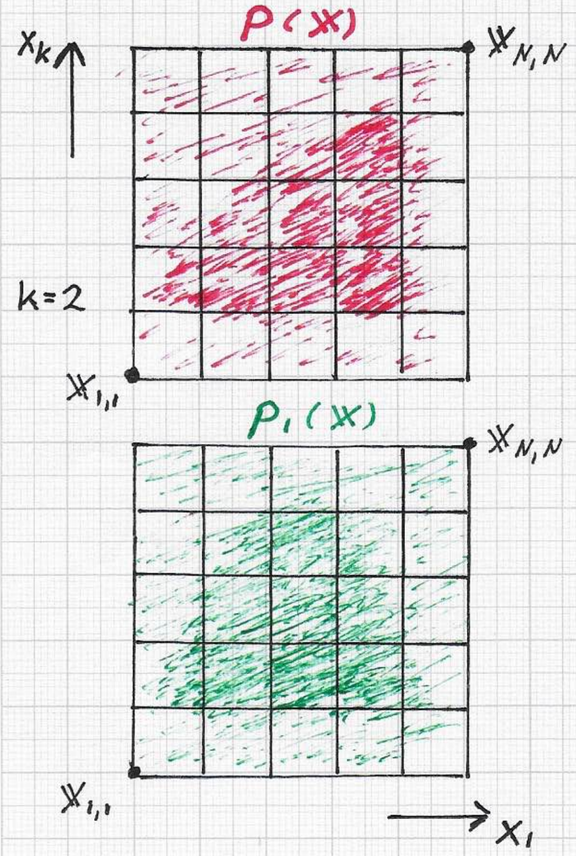
Stratoran

11/8/2020

■ DISTRIBUTION APPROXIMATION - Cont'd.

→ Considerations and Issues (Details):

5) Comparison of discretized NEW unclassified p-function, $p(\mathbb{X})$, and model p-function, $p_1(\mathbb{X})$, for example:



- $p(\mathbb{X})$ and $p_1(\mathbb{X})$ are represented by samples $p(\mathbb{X}_{1,\dots,1}), \dots, p(\mathbb{X}_{N,\dots,N})$ and $p_1(\mathbb{X}_{1,\dots,1}), \dots, p_1(\mathbb{X}_{N,\dots,N})$.
- The number of p-function samples is N^k.
- The number of features should be SMALL!
- Ex.: $k = 5$, $N = 32$ $\Rightarrow 32^5 = 2^{25}$ p-function samples
- Shaded versions of $p(\mathbb{X})$ and $p_1(\mathbb{X})$ to be compared
- Ex.: resolution of 3D image: 512^3; (average) number of voxels of an object to be classified: $64^3 = 2^{18}$

⇒ Root-mean-square (RMS) difference between $p(\mathbb{X})$, $p_1(\mathbb{X})$:

$$d = \sqrt{\sum_{\mathbf{i}} \left(p(\mathbb{X}_{\mathbf{i}}) - p_1(\mathbb{X}_{\mathbf{i}}) \right)^2},$$

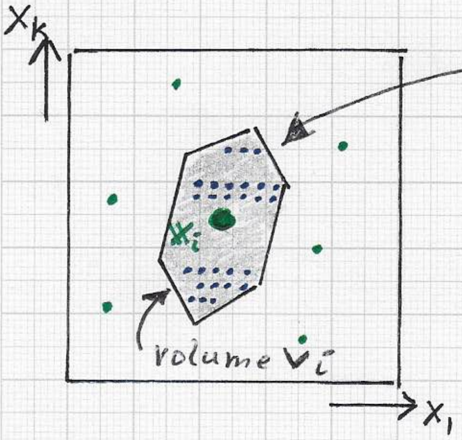
$\mathbf{i} = (1,\dots,1), \dots (N,\dots,N)$

$\Rightarrow n = 2^{18}$ = no. of sites used for DISCRETE SIBSON interpolation

▪ DISTRIBUTION APPROXIMATION – Cont'd.

➜ More Considerations...

6) Computation/estimation of k-dim. tiles (= hyper-volumes) for definition of ρ_i values associated with original feature points



tile of site $\ast_i$ (= •), tile hyper-volume = ?

⇒ Define a high-resolution sampling approach to estimate hyper-volume V_i; determine set of samples in $\ast$-space closest to $\ast_i$; assuming that this sample point set {⁝⁝⁝⁝} consists of centers of hyper-cubes in k-dim. feature space, add the hyper-cube volumes to estimate V_i.

•••••• = points closest to $\ast_i$; hypercube volume associated with ⊡ = c

$$\Rightarrow V_i = C \cdot c,$$

C = cardinality of set of all '•' inside $\ast_i$'s tile

⇒ $\ast_i$ has multiplicity m_i:

$$\rho_i = \frac{m_i}{V_i}$$

---

i) ρ_i-values for sites resulting from "training a high-quality" p-function can be pre-generated using high-res. sampling.

ii) ρ_i-values for sites used to define a p-function for an object in a NEW 3D scan must be computed RAPIDLY!

PARALLEL COMPUTATION

■ DISTRIBUTION APPROXIMATION – Cont'd.

➔ Considerations...

7) Resolutions are crucially important for storage/memory cost, computational/performance cost, and classification quality. These are possible resolution scenarios:

i) 3D image resolution: $512^3 = 2^{27}$

ii) Number of features: $5 = k$

iii) Number of voxels per segmented object: $64^3 = 2^{18}$

iv) Number of same-class objects (used for "training" the model): $16 = 2^4$

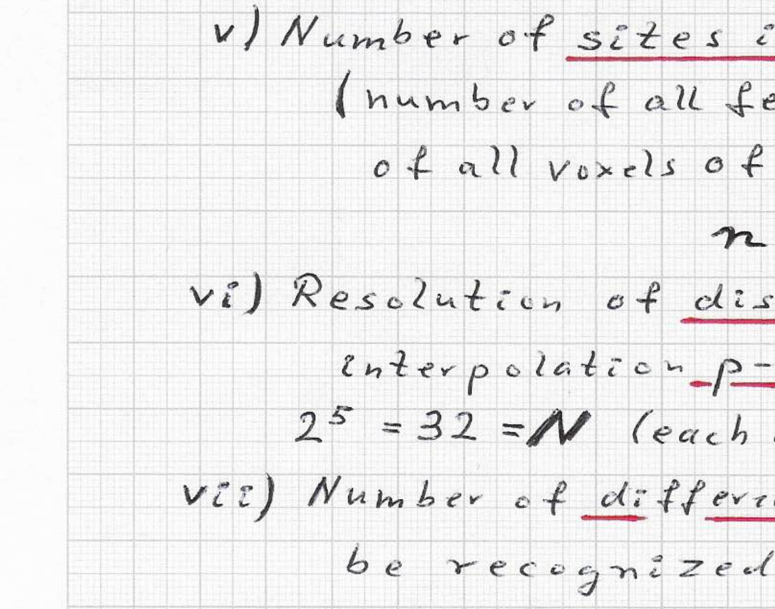
v) Number of sites in feature space (number of all feature points/vectors of all voxels of all same-class objects):

$$n = 16 \cdot 64^3 = 2^4 \cdot 2^{18} = 2^{22}$$

vi) Resolution of discretized Sibson interpolation p-function over k-dim. space: $2^5 = 32 = N$ (each dimension) $\Rightarrow N^k = 2^{25}$

vii) Number of different materials to be recognized: $16 = 2^4$

viii) Number of voxels of segmented object to be classified: $64^3 = 2^{18}$



● **Example:**

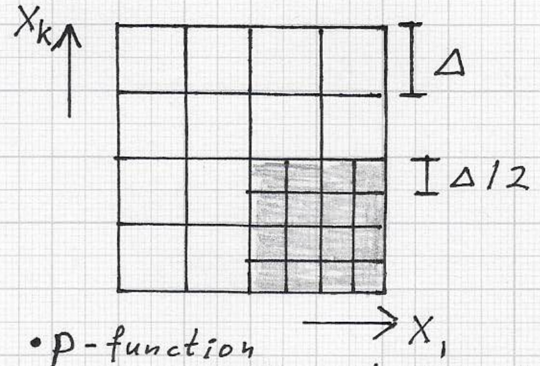
$\{x_1, x_2, \dots, x_{2^{18}}\}$ = single object's feature points in 5-dim. feature space; density values ρ_i used to compute p-values at $x_{i,j,k,l,m}$

Strateran

■ DISTRIBUTION APPROXIMATION – Cont'd.

→ Considerations...

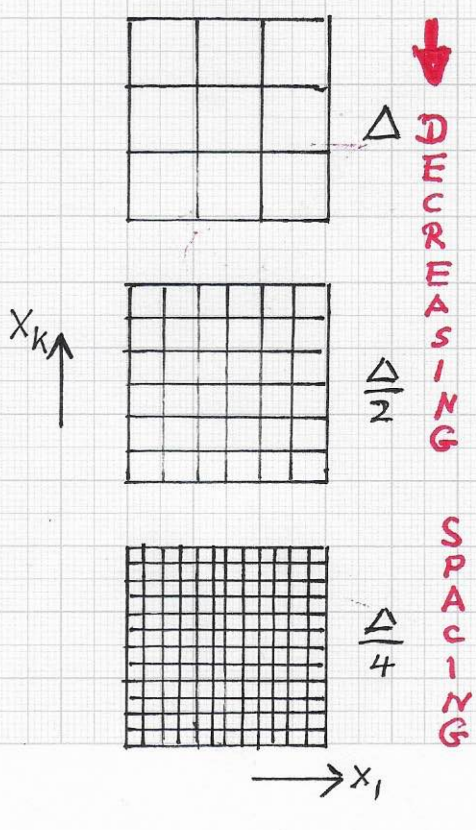
8) Adaptive grid spacing (= Locally varying resolution) can be used to compute higher-precision p-function values where needed:



• p-function grids – adaptive

• For example, use grid spacing $\Delta/2$ in regions of the p-function's domain where the p-function varies more, i.e., where more feature points/vectors are given. (The discrete Sibson interpolation step can incorporate adaptive spacing by using spacing-based weighted ρ_i-value additions.)

• p-function grids

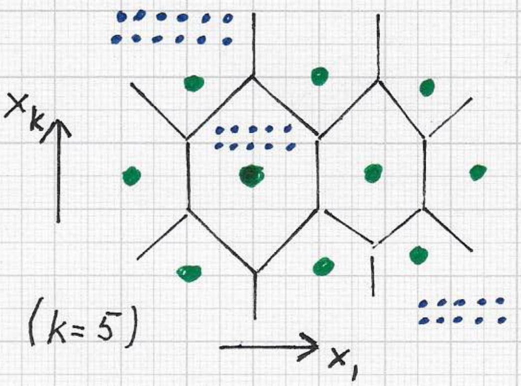


9) "Accelerated convergence" of discrete Sibson-based p-function via RICHARDSON extrapolation: spacing → 0. Using a measure for convergence of a gridded p-function, compute a sequence of such gridded functions using spacings Δ, $\Delta/2$, $\Delta/4$, ... until threshold is met.

■ DISTRIBUTION APPROXIMATION – Cont'd.

➡ Considerations...

10) Ratio of number of sites (tiles) of Sibson interpolation function and number of values of the p-function (evaluated on a uniform grid):



$\{\bullet\}$ = set of sites $= \{x_i\}$

$\{\cdot\}$ = p-function values

Example: (see p. 13)

i) no. sites for MODEL: 2^{22}

ii) no. sites for OBJECT to be classified: 2^{18}

iii) no. of p-function values: 2^{25}

• Goal: much larger number of p-function values than number of sites...!

$$\Rightarrow \text{Ratios } \frac{iii}{i} = \frac{2^{25}}{2^{22}} = 2^3 = 8$$

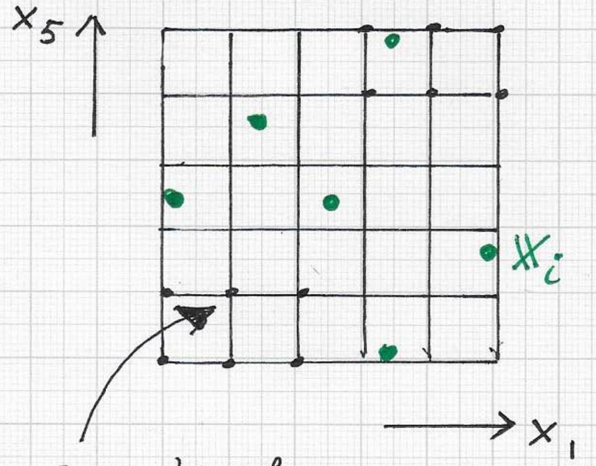
$$\frac{iii}{ii} = \frac{2^{25}}{2^{18}} = 2^7 = 128$$

• The feature space between sites (measured) and the feature space sampling density of the p-function (obtained via discrete Sibson interpolation) must be related such that desired classification performance is ensured.

DISTRIBUTION APPROXIMATION – Cont'd.

⇒ Considerations ...

III) Normalization and floating-point representation: (i) A "meaningful normalization" of the individual dimensions of feature points/vectors $\mathbb{X}$ is necessary. (ii) The feature space grid points used to sample the p-function have floating-point coordinates in feature space; the p-values are also represented as floating-point numbers. (iii) All given, measured sites $\mathbb{X}$ also use floating-point representation:



sample of p-function

→ original feature point/vector tuples:

• $= \mathbb{X}_i = (x_1^i, \ldots, x_5^i)$

→ "multi-dimensional" (5-dim.) array of uniformly sampled $\mathbb{X}$- and associated p-values:

• $= \mathbb{X}_{\mathbf{i}} = (x_1^{\mathbf{i}}, \ldots, x_5^{\mathbf{i}})$

with multi-index $\mathbf{i} = (i_1, i_2, i_3, i_4, i_5)$;

$p_{\mathbf{i}} = p_{i_1, i_2, i_3, i_4, i_5}$.

~ BH