PA170 Digital Geometry

Lecture 09: Content Measurement

Martin Maška (xmaska@fi.muni.cz)

Centre for Biomedical Image Analysis
Faculty of Informatics, Masaryk University, Brno

Autumn 2023

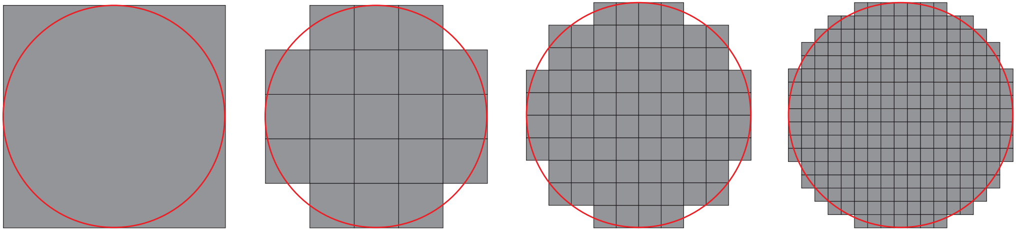
Motivation: Convergence of Estimators

- A real disk D of unit diameter has the area $\mathcal{A}(D) = \frac{\pi}{4}$ and the perimeter $\mathcal{P}(D) = \pi$
- The area of a digitized disk converges toward the area of the real disk with an increasing grid resolution h:

$$\lim_{h \to \infty} \mathcal{A}(dig_h(D)) = \mathcal{A}(D) = \frac{\pi}{4}$$

- The perimeter of a digitized disk does not converge toward the perimeter of the real disk:

$$\lim_{h \to \infty} \mathcal{P}(dig_h(D)) = 4$$



$h = 1$ $h = 5$ $h = 10$ $h = 17$

Multigrid Convergence

- Let $\mathbb{F}$ be a family of sets S in $\mathbb{R}^n$, $dig_h(S)$ be a digitization of S on a grid of resolution h, and Q be a property (e.g., area, perimeter, or length) defined for all $S \in \mathbb{F}$
- An estimator E_Q is called multigrid convergent for $\mathbb{F}$ and for dig_h iff, for any $S \in \mathbb{F}$, there is a grid resolution $h_S > 0$ such that the estimated value $E_Q(dig_h(S))$ is defined for any grid resolution $h \geq h_S$, and

$$|E_Q(dig_h(S)) - Q(S)| \leq \kappa(h)$$

where κ is a speed of convergence function that converges toward zero as $h \to \infty$

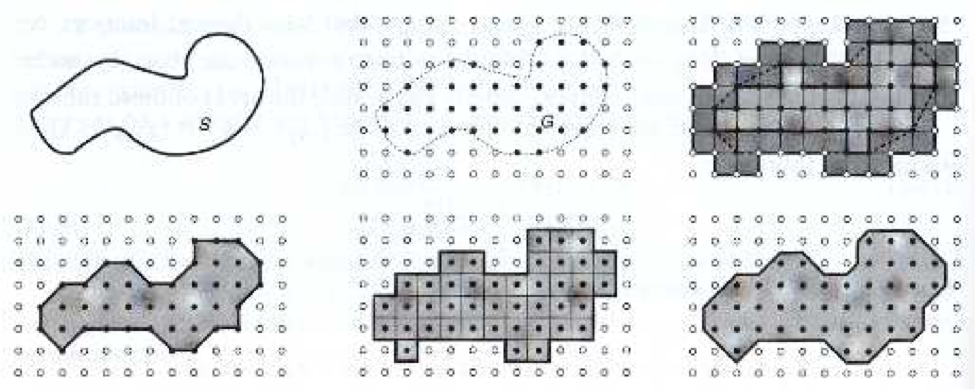
Examples of theoretical results (for any grid resolution $h > 0$ and Gauss digitization G_h)

- For any planar convex set S, $|\mathcal{A}(G_h(S)) - \mathcal{A}(S)| = \mathcal{O}(h^{-1})$ [Gauss & Dirichlet]
- For any centered disk D, $|\mathcal{A}(G_h(D)) - \mathcal{A}(D)| = \Omega(h^{-1.5})$ [Hardy 1913]
- For any planar convex 3-smooth set S, $|\mathcal{A}(G_h(S)) - \mathcal{A}(S)| = \mathcal{O}(h^{-\frac{100}{73}} \cdot (\log h)^{\frac{315}{146}})$ [Huxley 1993]

AREA ESTIMATION

Area Estimators

- Let $S \subseteq \mathbb{R}^2$ be a planar compact set and $\mathcal{A}(S)$ be its true area
- The area of the Gauss digitization $G_h(S)$ converges toward $\mathcal{A}(S)$
- The area of the inner and outer Jordan digitizations $J_h^-(S)$ and $J_h^+(S)$, respectively, converges toward $\mathcal{A}(S)$ too
- Therefore, the area of any digitization between $J_h^-(S)$ and $J_h^+(S)$ also converges toward $\mathcal{A}(S)$

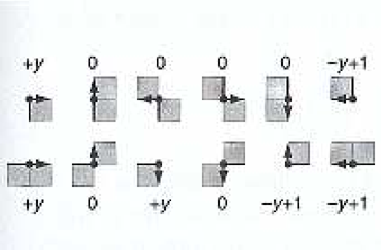


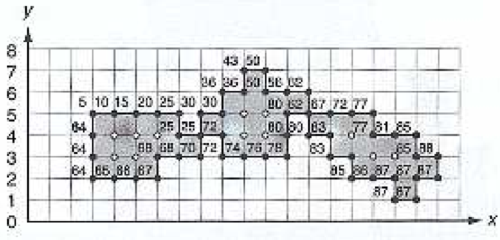
- The area $\mathcal{A}(\Pi)$ of an isothetic grid polygon Π can be calculated as

$$\mathcal{A}(\Pi) = \frac{1}{h^2} \cdot \left(\alpha_0 - \frac{L}{2} - 1\right)$$

where $h > 0$ is grid resolution, α_0 is the number of grid points in Π, and L is the total length of its frontier

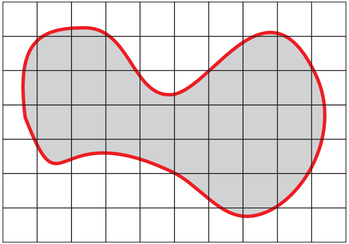
- Both L and α_0 can easily be calculated during border tracing. In particular, α_0 can be calculated using discrete column-wise integration:
  1. $\alpha_0 = 0$
  2. $\alpha_0 = \alpha_0 + y$ for all grid points (x, y) at the upper end of a vertical run of object grid points
  3. $\alpha_0 = \alpha_0 - y + 1$ for all grid points (x, y) at the bottom end of such a run



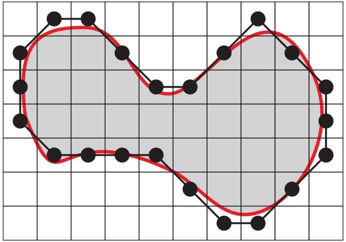


LENGTH ESTIMATION

- The frontier of a simply connected, planar compact set S is a simple, rectifiable curve $\gamma : [0, 1] \to \mathbb{R}^2$
- Three possible digitizations of γ are as follows:
  - **[A]** A cyclic ordered sequence $\rho_h(\gamma)$ of grid points derived from the grid-intersection digitization of γ in $\mathbb{Z}_h^2$
  - **[B]** A cyclic ordered sequence of grid vertices of 2-cells on the frontier of the Gauss digitization $G_h(S)$ of S
  - **[C]** The closed difference set between the outer and inner Jordan digitizations (i.e., $M = (J_h^+(S) \setminus J_h^-(S))^\bullet$)



Set S and its topological frontier γ



Cyclic ordered sequence $\rho_h(\gamma)$

Local Estimators ([A])

- The true curve length $\mathcal{L}(\gamma)$ is approximated as a weighted sum of different types of steps in $\rho_h(\gamma)$, which is easy to implement and unique, but not multigrid convergent
- Let n_i and n_d be the number of isothetic and diagonal steps, respectively, and n_c be the number of transitions between these two types of steps in $\rho_h(\gamma)$
- The geometric length estimator approximates $\mathcal{L}(\gamma)$ as

$$\mathcal{L}_{GEOM}(\rho_h(\gamma)) = \frac{1}{h} \cdot (n_i + \sqrt{2} \cdot n_d)$$

- The best linear unbiased estimator minimizes the mean square error between the estimated and true curve length:

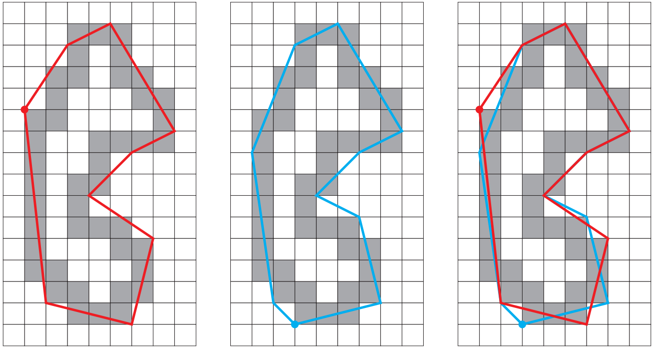
$$\mathcal{L}_{BLUE}(\rho_h(\gamma)) = \frac{1}{h} \cdot (0.948 \cdot n_i + 1.343 \cdot n_d)$$

- The cornercount estimator approximates $\mathcal{L}(\gamma)$ as

$$\mathcal{L}_{COC}(\rho_h(\gamma)) = \frac{1}{h} \cdot (0.980 \cdot n_i + 1.406 \cdot n_d - 0.091 \cdot n_c)$$

DSS-Based Estimators ([B])

- The true curve length $\mathcal{L}(\gamma)$ is approximated by integrating lengths of the maximum-length digital straight line segments in the digitization of γ, which is easy to implement and multigrid convergent, but not unique
- The basic DSS-based estimator calculates $\mathcal{L}_{DSS}(dig_h(\gamma))$ as the length of the resulting polygon (or of the polygonal arc in case of open curves)
- The most probable original length estimator calculates $\mathcal{L}_{MPO}(dig_h(\gamma))$ by replacing the real DSS lengths in $\mathcal{L}_{DSS}(dig_h(\gamma))$ with $\frac{n}{h} \cdot \sqrt{1 + a_h^2}$ where n is the length of the binary-word representation of a particular DSS and a_h is the estimation of its slope



Tangent-Based Estimator: Preliminaries

- Let $\gamma(t)$ be a parametrized curve (i.e., $\gamma(t) = (x(t), y(t))$, $a \leq t \leq b$)
- Apart from uniquely determining the geometric location of all the curve points, a parametrization provides information about the curve orientation and its speed $v(t)$:

$$v(t) = \|\dot{\gamma}(t)\|_2 = \|(\dot{x}(t), \dot{y}(t))\|_2$$

where

$$\dot{x}(t) = \frac{dx(t)}{dt} \quad \text{and} \quad \dot{y}(t) = \frac{dy(t)}{dt}$$

- If γ is rectifiable, its length $\mathcal{L}(\gamma)$ is given as

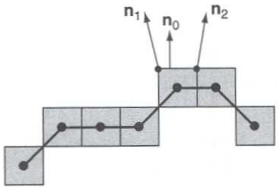
$$\mathcal{L}(\gamma) = \int_a^b v(t)dt = \int_a^b \|(\dot{x}(t), \dot{y}(t))\|_2 dt = \int_a^b \sqrt{(\dot{x}(t))^2 + (\dot{y}(t))^2} dt$$

Tangent-Based Estimator ([B])

- The true curve length $\mathcal{L}(\gamma)$ is approximated by integrating $\|(\dot{x}, \dot{y})\|_2$ along $dig_h(\gamma)$, which is multigrid convergent and unique, but substantially slower than local and DSS-based estimators due to the cost associated with the estimation of normals
- By tracing the ordered sequence of 0-cells and estimating digitized curve normals at these locations using a DSS algorithm, the tangent-based estimator approximates $\mathcal{L}(\gamma)$ as

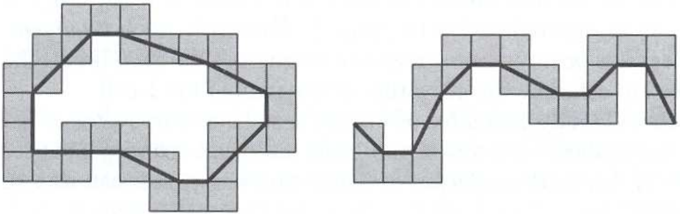
$$\mathcal{L}_{TAN}(dig_h(\gamma)) = \sum_{p \in A} \frac{\mathbf{n_1}(p) + \mathbf{n_2}(p)}{2} \cdot \mathbf{n}_0(p)$$

where A is the set of all frontier 1-cells of $dig_h(\gamma)$, $\mathbf{n_0}(p)$ is the unit normal to $p \in A$, and $\mathbf{n_1}(p)$ and $\mathbf{n_2}(p)$ are the estimated normals at the endpoints of p



MLP-Based Estimators ([C])

- In case of closed curves, MLP is a minimum-length polygon that circumscribes the inner frontier of *M* and is in the interior of its outer frontier
- In case of open curves, MLP is a minimum-length polygonal arc that is incident with all 2-cells in M
- The MLP-based estimator calculates $\mathcal{L}_{MLP}(dig_h(\gamma))$ as the length of the resulting polygon (or the polygonal arc in case of open curves), which is multigrid convergent and unique, but slower than local and DSS-based estimators due to the cost associated with the MLP construction



Comparison of Length Estimators: Main Observations

- A comparative study of the introduced length estimators conducted in [Coeurjolly & Klette, 2004]
- The analyzed dataset contained convex as well as nonconvex shapes, digitized on grids of sizes between 30×30 and 1000×1000 grid points

Main Observations

- All the evaluated estimators converge, but the local ones toward false values with relative errors about 2 % at maximum grid resolution
- All the evaluated estimators run in linear time, but TAN is roughly three times slower than its competitors at maximum grid resolution
- All the evaluated estimators but the local ones are nearly orientation-independent, with relative errors up to 2 %. The relative error committed by BLUE is from 4 % to 12 % for a square rotated between 30 and 60 degrees

Length Estimators: Summary

Method	Multigrid	Discrete	Unique
GEOM	No	Possibly	Yes
BLUE	No	Possibly	Yes
COC	No	Possibly	Yes
DSS	Yes	Yes	No
MPO	Yes	Yes	No
TAN	Yes	Yes	Yes
MLP	Yes	Yes	Yes

Multigrid Is the estimator multigrid convergent at least for convex curves?

Discrete Does the core of the estimation algorithm deal only with integers?

Unique Is the result independent of initialization?

Take-Home Messages

- The area of Gauss as well as inner and outer Jordan digitizations of planar compact sets converge to true areas of these sets
- Local length estimators are fast and unique, but not multigrid convergent
- DSS-based and MPO-based estimators are multigrid convergent, but not unique
- MLP-based and TAN-based estimators are multigrid convergent and unique at the expense of their speed