Lattice Computing : a partial history

Manuel Graña Romay

Grupo de Inteligencia Computacional (GIC); UPV/EHU; www.ehu.es/ccwintco

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 777720

Localization

The GIC belongs to the University of the Basque Country (UPV/EHU)

Located in Donostia-San Sebastian, Spain, Europe

Summary of group works: www.ehu.es/ccwintco > Historial de grupo

Contents

- Introductory ideas and history
- Filtering
  - Fuzzy Mathematical Morphology
  - Multivariate Mathematical Morphology
- Classification
  - Fuzzy ART
  - Max-min classifiers
  - Fuzzy Lattice Neurocomputing
- Associative Morphological Memories
- Conclusions and the future

Introduction

- Lattice & Computing: suggestions
  - Parallel Computing
    - Computing elements arranged as a lattice, each having channels to neighbouring elements
  - A representation of particle interaction in physics in crystal solids
  - A spatial discretization for application of finite element methods

Introduction

- Lattice computing assumes that the basic computing structure is a lattice.

- A lattice $(L,\vee,\wedge)$ is a Poset $(L,\leq)$ any two of whose elements have
  - a supremum, denoted by $x \vee y$
  - an infimum, denoted by $x \wedge y$

Introduction

- Poset

A *partially-ordered set*, briefly **poset** $(\mathcal{P}, \leq)$, is a set $\mathcal{P}$ in which a binary relation $\leq$ is defined that is a *partial ordering*, i.e., satisfies the following three properties for all $X, Y, Z \in \mathcal{P}$:

(P1). $X \leq X$ (reflexive)
(P2). $X \leq Y$ and $Y \leq X$ imply $X = Y$ (antisymmetric)
(P3). $X \leq Y$ and $Y \leq Z$ imply $X \leq Z$ (transitive)

Introduction

- Computational paradigm shift (Ritter)
  - Traditional Artificial Neural Networks are defined on the ring $(\mathbb{R}, +, \times)$

$$\tau_j(\mathbf{x}) = \sum_{i=1} x_i w_{ij} - \theta_j$$

  - Lattice ANN work on the semi-rings

$$(\mathbb{R}_{-\infty}, \vee, +) \text{ or } (\mathbb{R}_{\infty}, \wedge, +)$$

$$\tau_j(\mathbf{x}) = p_j \bigvee_{i=1}^{n} r_{ij}(x_i + w_{ij}) \qquad \tau_j(\mathbf{x}) = p_j \bigwedge_{i=1}^{n} r_{ij}(x_i + w_{ij})$$

Introduction

- Biological justification (Ritter)
  - Dendrites account for 50% of brain mass
  - Dendrite computation is more akin to AND, XOR, NOT logical operations

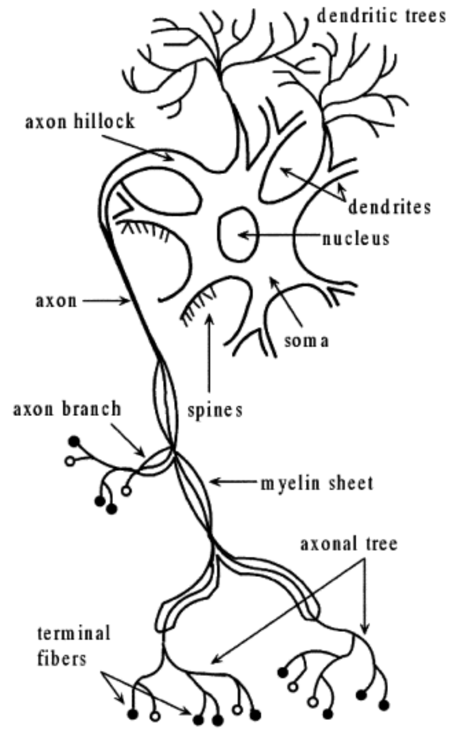


Fig. 1. Diagram of a neuron cell showing dendrites, dendritic trees, axon branches, and terminal branches. Excitatory and inhibitory inputs are indicated, respectively, by black small disks (•) and small circles (○).

Introduction

- Mathematical morphology for image processing is also a lattice paradigm shift from linear processing (Maragos)
  - Linear translation-invariant (LTI) operators are uniquely represented by linear convolution with the impulse response
  - Erosion (Dilation) translation invariant (ETI(DTI)) operators are uniquely represented by inf-(sup) convolution with the impulse response

Introduction

$$\psi \text{ is LTI} \Leftrightarrow \psi(F)(x) = (F * H)(x) = \sum_{y} F(y) H(x-y)$$

DTI $$(F \circledast H)(x) \triangleq \bigvee_{y \in \mathbb{E}} F(y) \star H(x-y)$$

ETI $$(F \circledast' H')(x) \triangleq \bigwedge_{y \in \mathbb{E}} F(y) \star' H'(x-y)$$

Introduction

- Kinds of processes in Artificial Intelligence
  - Filtering

$$\psi : R^N \rightarrow R^N$$

  - Dimension reduction

$$\psi : R^N \rightarrow R^d ; d << N$$

  - Classification (supervised, unsupervised)

$$\psi : R^N \rightarrow \Omega; \quad \Omega = \{\omega_1, \ldots, \omega_c\}$$

Introduction

- Lattice Computing approaches
  - Filtering: Mathematical Morphology
  - Dimension reduction: ? ? ? ? ? ? ?
  - Classification- recognition
    - Fuzzy systems
    - Artificial Neural Networks
  - Specific processes
    - Target Localization in images
    - Endmember induction in hyperspectral images

Introduction

- The learning problem
  - Gradient descent schemas need to compute derivatives :
    - derivatives of sup, inf functions are not defined.
  - Heuristic growing produces overfitting (category explosion) and there is no proof of convergence.
  - Random search algorithms are computationally expensive.

Some historical landmarks

- 1979
  - R. Cuninghame-Green: Minimax Algebra
- 1982
  - J. Serra: Image Analysis and Mathematical Morphology
- 1991
  - Carpenter, Grossberg: Fuzzy-ART
- 1992
  - Simpson: Min-max Neural Networks
  - Pedrycz: Relational System Learning
- 1995
  - Yang, Maragos: Min-max Classifiers
- 1998
  - Ritter, Sussner: Morphological Associative Memories
  - Gader: Shared-weight Morphological Neural Networks
- 2000
  - Kaburlassos, Petridis: Fuzzy Lattice Neurocomputing
- 2003
  - Ritter: Dendritic Computing
- 2005
  - Kaburlassos: Towards a unified modeling and knowledge representation based on Lattice Therory
  - Maragos: Lattice image processing: a unification of morphological and Fuzzy algebraic systems
- 2007
  - Kaburlassos, Ritter: Computational Intelligence based on Lattice Theory

Contents

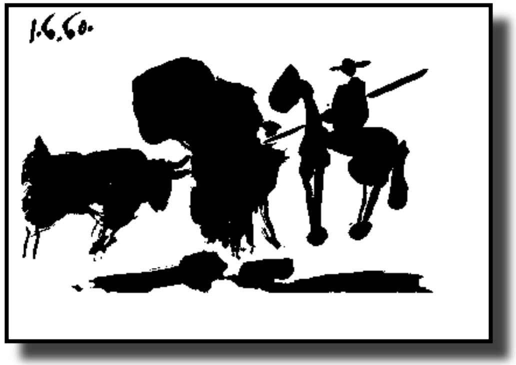
- Introductory ideas and history
- Filtering
  - Fuzzy Mathematical Morphology
  - Multivariate Mathematical Morphology
- Classification
  - Fuzzy ART
  - Max-min classifiers
  - Fuzzy Lattice Neurocomputing
- Associative Morphological Memories
- Conclusions and the future

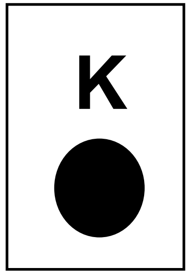
Lattice Image Processing: A Unification of Morphological and Fuzzy Algebraic Systems

P. Maragos

Journal of Mathematical Imaging and Vision 22: 333–353, 2005

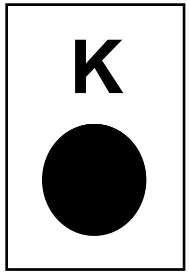
Binary erosion.-





Structuring element

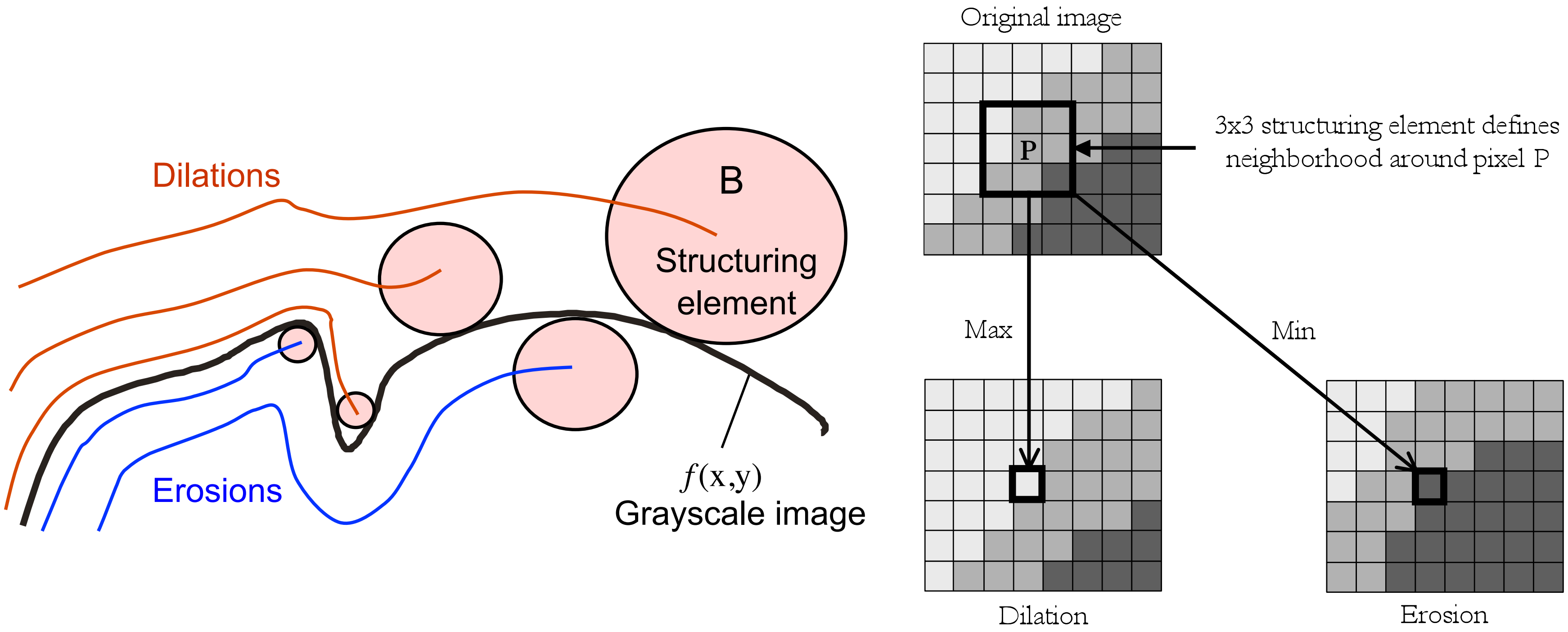
Binary dilation.-



Structuring element

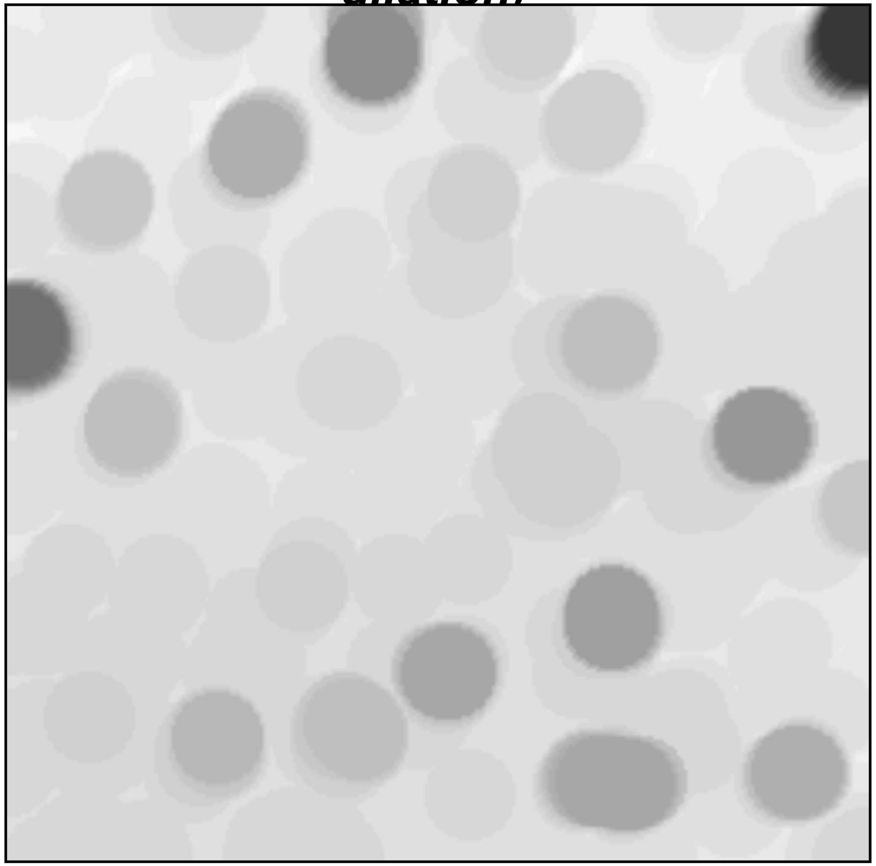
Greyscale Mathematical Morphology

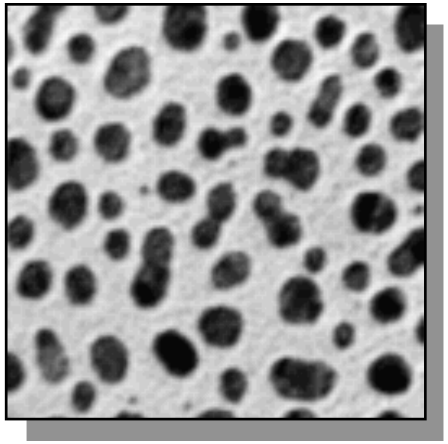
- *Grayscale* morphology relies on a *partial* ordering relation between image pixels.

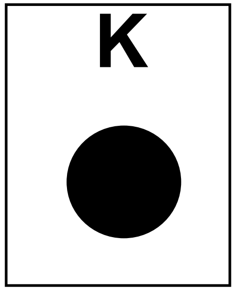


- Opening and closing: shape-preserving operators.
- Excellent filtering properties:

Morphological opening (erosion + dilation)







Structuring element

Starting point

- Design of new filters: generalized opening and closing
- Works on the lattice of functions

$$\mathcal{S} = \mathbb{V}^E \qquad F : E \to \mathbb{V}$$

$$F \leq G \Leftrightarrow F(x) \leq G(x) \;\; \forall x \in E$$

Inherited partial order

$$\Big(\bigvee_{i \in J} F_i\Big)(x) \triangleq \bigvee_{i \in J} F_i(x), \quad x \in E$$

$$\Big(\bigwedge_{i \in J} F_i\Big)(x) \triangleq \bigwedge_{i \in J} F_i(x), \quad x \in E$$

Inherited supremum and infimum

Increasing operators

δ is **dilation** iff $\delta(\bigvee_{i\in J} X_i) = \bigvee_{i\in J} \delta(X_i)$

ε is **erosion** iff $\varepsilon(\bigwedge_{i\in J} X_i) = \bigwedge_{i\in J} \varepsilon(X_i)$

α is **opening** iff α is increasing, idempotent & anti-extensive

β is **closing** iff β is increasing, idempotent & extensive

Adjunction

- The operator pair (ε, δ) is an **adjunction** if

$$\delta(X) \leq Y \Leftrightarrow X \leq \varepsilon(Y) \quad \forall X, Y \in \mathcal{L}$$

- An adjunction defines a pair of morphological filters

$\delta\varepsilon$ *is an opening, and* $\varepsilon\delta$ *is a closing.*

Signal processing

- Algebraic structure of the scalars:

$$(\mathbb{V}, \vee, \wedge, \star, \star')$$

- Complete lattice-ordered double monoid
  - Addition $\vee$
  - Dual addition $\wedge$
  - Multiplication $\star$
  - Dual multiplication $\star'$

Signal processing

- The space of signals is a function lattice

$$\mathcal{S} = \mathrm{Fun}(E, \mathbb{V})$$

- It inherits the clodum structure of the scalars, with appropriate natural definitions of addition and multiplication

Parallelism to linear processing

- Representation of a signal as a supremum (infimum) of translated impulses

$$F(x) = \bigvee_{y \in E} F(y) \star q_y(x) = \bigwedge_{y \in E} F(y) \star' q'_y(x)$$

- Linear superposition principle

$$\psi\left(\sum_{i\in J} a_i \cdot F_i\right) = \sum_{i\in J} a_i \cdot \psi(F_i)$$

- Nonlinear superposition principle

$$\delta\left(\bigvee_{i\in J} c_i \star F_i\right) = \bigvee_{i\in J} c_i \star \delta(F_i),$$

- Translation invariant operator: commutes with all translations

$$\tau \in \mathbb{T}; \text{i.e. } \psi\tau = \tau\psi.$$

- Nonlinear convolutions define the effect of Erosion and Dilation translation invariant systems

$$\psi \text{ is LTI} \Leftrightarrow \psi(F)(x) = (F * H)(x) = \sum_{y} F(y) H(x-y)$$

DTI $$(F \circledstar H)(x) \triangleq \bigvee_{y \in \mathbb{E}} F(y) \star H(x-y)$$

ETI $$(F \circledstar' H')(x) \triangleq \bigwedge_{y \in \mathbb{E}} F(y) \star' H'(x-y)$$

Generalized convolution adjunctions

- using scalar adjunctions $(\overleftarrow{\lambda}_{H(x-y)}, \lambda_{H(x-y)})$
- It is possible to obtain the adjoint operator

$$\Delta_H(F)(x) = \bigvee_{y \in \mathbb{E}} F(y) \star H(x-y) \quad = \bigvee_{y \in \mathbb{E}} \lambda_{H(x-y)}(F(y))$$

- Which looks like a correlation

$$\mathcal{E}_H(G)(x) = \bigwedge_{y \in \mathbb{E}} G(y) \star [H(y-x)]^*$$

Lattice operators using fuzzy norms

- Fuzzy intersection norm --> scalar dilation

$$T: [0, 1]^2 \rightarrow [0, 1]$$

F1. $T(a, 1) = a$ and $T(a, 0) = 0$

F2. $T(a, T(b, c)) = T(T(a, b), c)$ (associativity).

F3. $T(a, b) = T(b, a)$ (commutativity).

F4. $b \leq c \Rightarrow T(a, b) \leq T(a, c)$ (increasing).

F5. T is a continuous function.

- Fuzzy union norm --> scalar erosion

$$U : [0, 1]^2 \rightarrow [0, 1]$$

$$\text{F1}'.\ U(a, 0) = a \text{ and } U(a, 1) = 1.$$

- Translations under the fuzzy framework

$$\mathcal{S} = \mathrm{Fun}(\mathbb{E}, [0, 1])$$

$$\tau_{h,v}(f)(x) = T(f(x - y), v)$$

$$\tau'_{h,v}(f)(x) = U(f(x - y), v)$$

$$(h, v) \in \mathbb{E} \times [0, 1]$$

- Signal representation with fuzzy translations

$$f(x) = \bigvee_y T[q(x-y), f(y)]$$
$$= \bigwedge_y U[q'(x-y), f(y)]$$

$$q(x) \triangleq \begin{cases} 1, & x = \vec{0} \\ 0, & x \neq \vec{0} \end{cases}, \quad q'(x) \triangleq \begin{cases} 0, & x = \vec{0} \\ 1, & x \neq \vec{0} \end{cases}$$

- Translation invariant signal fuzzy **dilations** and **erosions** with sup-*T* and inf-*U* convolutions

$$(f \bigcirc_T g)(x) \triangleq \bigvee_y T[g(x-y), f(y)],$$

$$(f \bigcirc'_U g)(x) \triangleq \bigwedge_y U[g(x-y), f(y)]$$

- Fuzzy dilation adjoint $\Delta_{H,T}(F)(x) \triangleq (F \bigcirc_T H)(x)$

$$\mathcal{E}_{H,\Omega}(G)(y) \triangleq \bigwedge_{x \in \mathbb{E}} \Omega[H(x-y), G(x)]$$

where $\Omega[H(x-y), G(x)]$ is actually the adjoint of the fuzzy T-norm:

$$T(a, v) \leq w \Leftrightarrow v \leq \Omega(a, w)$$

$$\Omega(a, w) = \sup\{v \in [0, 1] : T(a, v) \leq w\}$$

Example norms

Fuzzy intersection norm

$$\text{Min}: T_1(a,v) = \min(a,v)$$
$$\text{Product}: \quad T_2(a,v) = a \cdot v$$
$$\text{Yager}: T_3(a,v) = 1 - \big(1 \wedge [(1-v)^p + (1-a)^p]^{1/p}\big), \; p > 0.$$

Adjoint norm

$$\Omega_1(a,w) = \begin{cases} w, & w < a \\ 1, & w \geq a \end{cases}$$
$$\Omega_2(a,w) = \begin{cases} \min(w/a, 1), & a > 0 \\ 1, & a = 0 \end{cases}$$
$$\Omega_3(a,w) = \begin{cases} 1 - [(1-w)^p - (1-a)^p]^{1/p}, & w < a \\ 1, & w \geq a \end{cases}$$

Results

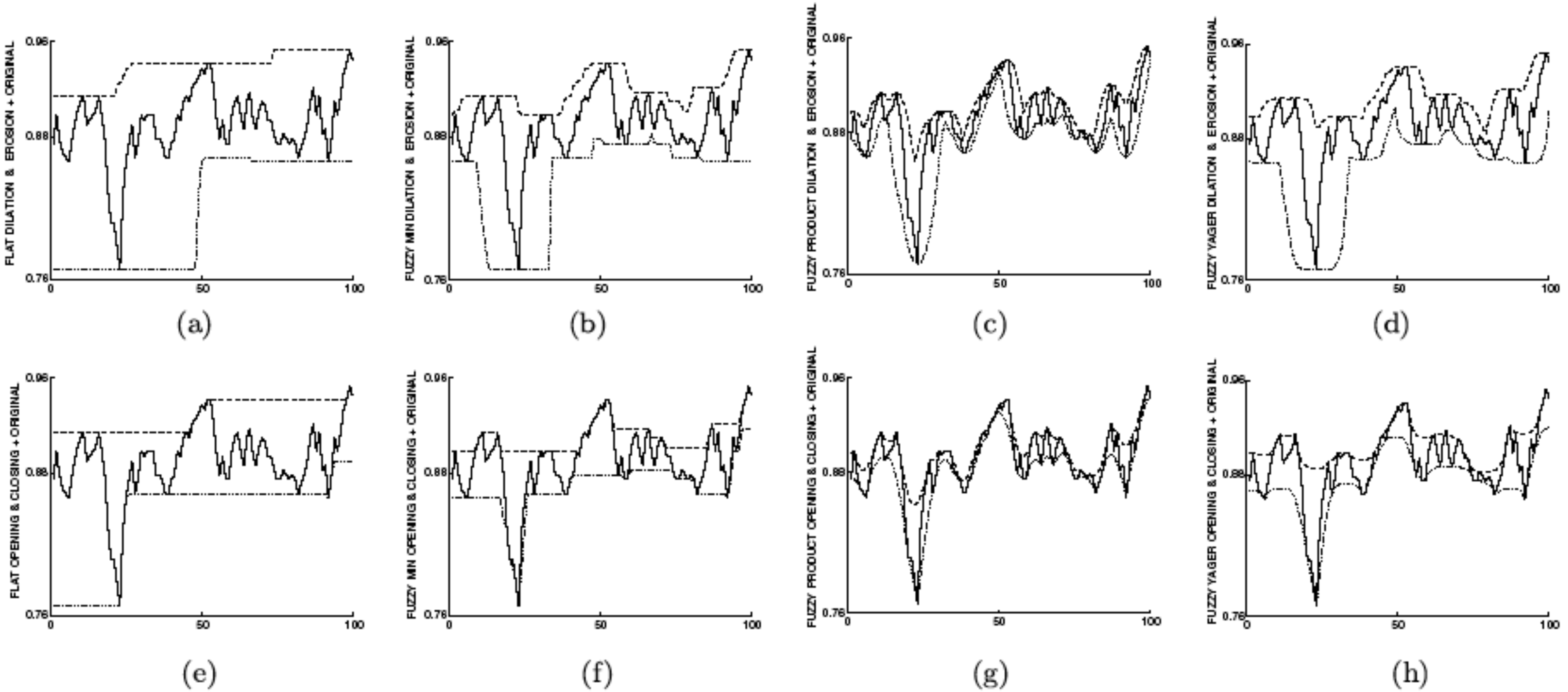
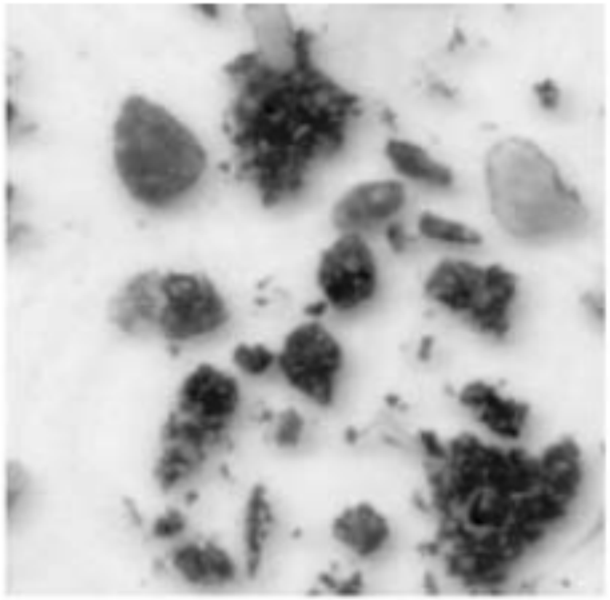


Figure 1. Comparison of 1D basic morphological and lattice-fuzzy signal operators. Rows 1 and 2, left to right: flat, minimum, product, Yager. Row 1: original signal (solid line), dilation (dashed line), erosion (dotted line). Row 2: closing (dashed line), opening (dotted line). Courtesy of [27].



(a)

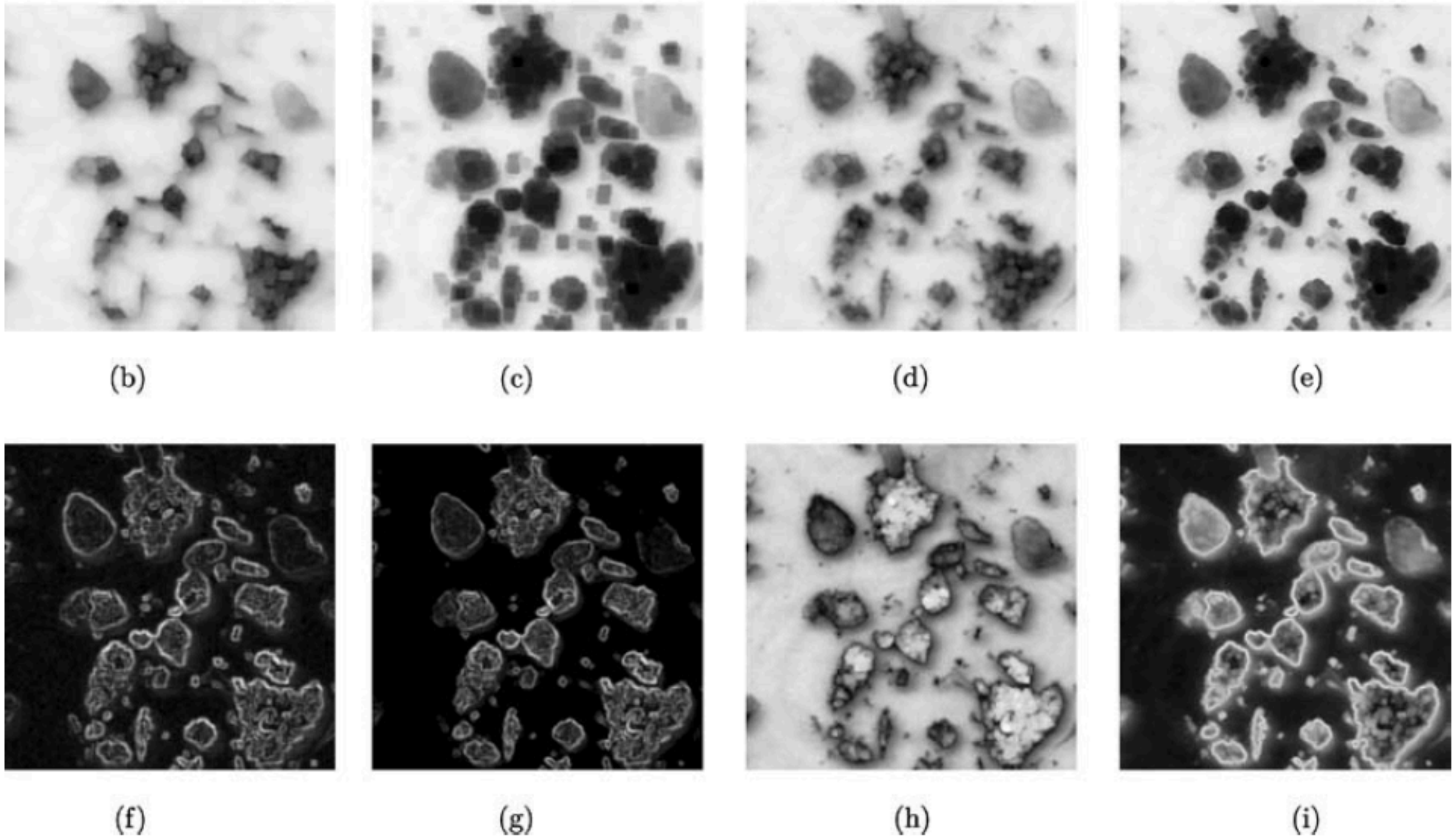


Figure 2. (a) Original image F. (b) Morphological flat dilation $F \oplus B$. (c) Morphological flat erosion $F \ominus B$. (d) Fuzzy dilation $\delta(F)$. (c) Fuzzy erosion $\mathcal{E}(F)$. (f) Morphological gradient $F \oplus B - F \ominus B$. (g) $\delta(F) - \mathcal{E}(F)$. (h) Fuzzy min gradient $\min[\delta(F), 1 - \mathcal{E}(F)]$. (i) Fuzzy max gradient $\max[\delta(F), 1 - \mathcal{E}(F)]$. Courtesy of [27].

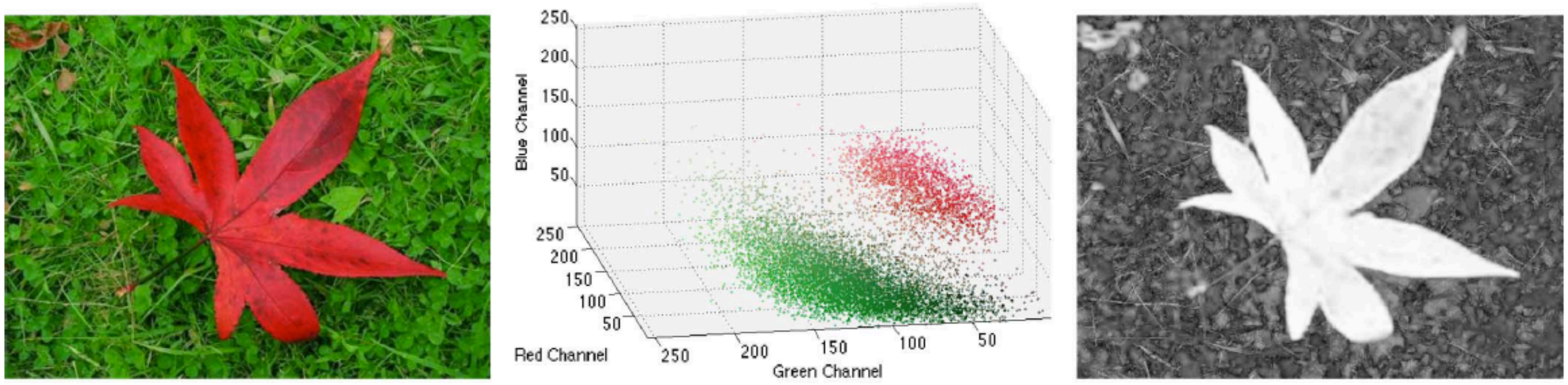
Random Projection Depth for MultivariateMathematical Morphology

Santiago Velasco-Forero, and Jesús Angulo

IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, VOL. 6, NO. 7, NOVEMBER 2012

- Multivariate signal (image) orders
  - Marginal: each channel separately
  - Conditional: lexicographic total order
  - Reduced: induced by a map into scalar
  - P-order: induced by partitioning of the vector sample into groups

- Depth functions assign to each point its degree of centrality with respect to a data cloud or a probability distribution: center-outward ordering of point



(a) (b) (c)

Fig. 1. The proposed ordering for a given multivariate image (a) is based on the information contained in its spectral representation (b). Projection depth function (c) detects the intrinsic dichotomy background and foreground of the original image. Total ordering for morphological transformations is defined as follows: $\mathbf{x}_1 < \mathbf{x}_2 \Leftrightarrow PD(\mathbf{x}_1; \mathbf{I}) < PD(\mathbf{x}_2; \mathbf{I})$.

- Projection Depth Function

Definition 2: [18] *The* projection depth function *for a vector* $\mathbf{x}$ *according with a data cloud* $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$ as follows,

$$PD(\mathbf{x};\mathbf{X}) = \sup_{\mathbf{u}\in\mathbb{S}^{d-1}} \frac{\left|\mathbf{u}^T\mathbf{x} - MED(\mathbf{u}^T\mathbf{X})\right|}{MAD(\mathbf{u}^T\mathbf{X})} \qquad (2)$$

  - MED: median; MAD median absolute deviation

$$PD(\mathbf{x};k,\mathbf{X}) = \max_{\mathbf{u}\in\mathbb{U}} \frac{\left|\mathbf{u}^T\mathbf{x} - \mathrm{MED}(\mathbf{u}^T\mathbf{X})\right|}{\mathrm{MAD}(\mathbf{u}^T\mathbf{X})} \qquad (3)$$

where $\mathbb{U} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ with $\mathbf{u}_i \in \mathbb{S}^{d-1}$. Clearly, if $k \to \infty$ then $PD(\mathbf{x};k,\mathbf{X}) \to PD(\mathbf{x};\mathbf{X})$.

Stochastic finite approximation

- h-ordering

let $h : \mathrm{R} \rightarrow \mathcal{L}$ be a surjective mapping.

$$\mathrm{r} =_h \mathrm{r}' \Leftrightarrow h(\mathrm{r}) = h(\mathrm{r}') \; \forall \mathrm{r}, \mathrm{r}'.$$

refer by $\leq_h$ the *h-ordering*

$$\mathrm{r} \leq_h \mathrm{r}' \Leftrightarrow h(I) \leq h(\mathrm{r}'), \quad \forall \mathrm{r}, \mathrm{r}' \in R$$

$h^{\leftarrow} : \mathcal{L} \rightarrow \mathrm{R}$ semi-inverse of h.

$$hh^{\leftarrow}(r) = r, \text{ for } r \in \mathcal{L}.$$

the pair (ε, δ) is called an *h-adjunction*.

$$\varepsilon, \delta \; : \mathrm{R} \to \mathrm{R}$$

$$\delta(\mathrm{r}) \leq_h \mathrm{r}' \Leftrightarrow \mathrm{r} \leq_h \varepsilon(\mathrm{r}'), \quad \forall \mathrm{r}, \mathrm{r}' \in \mathrm{R}$$

Moreover, let (ε, δ)

be h-increasing mappings on R, and let $\varepsilon \mapsto^h \widetilde{\varepsilon}$, $\delta \mapsto^h \widetilde{\delta}$. Then (ε, δ) is an h-adjunction on R if and only if $(\widetilde{\varepsilon}, \widetilde{\delta})$ is an adjunction on the lattice $\mathcal{L}$. Therefore a mapping δ (resp. ε) on R is called *h-dilation* (resp. *h-erosion*) if $\widetilde{\delta}$ (resp. $\widetilde{\varepsilon}$) is a dilation (resp. erosion) on $\mathcal{L}$.

$$\gamma = \delta\varepsilon \leq_h \mathrm{id} \leq_h \varphi = \varepsilon\delta.$$

h-opening &
h-closing

Given a multivariate vector image $\mathbf{I} \in \mathcal{F}(\mathbb{E}, \mathbb{F})$, its *h-depth mapping* is defined as

$$h_{\mathbf{I}}(\mathbf{x}) = PD(\mathbf{x}; \mathbf{X}_{\mathbf{I}}) \tag{12}$$

$$\varepsilon_{\mathrm{SE},h_{\mathbf{I}}}(\mathbf{I})(x) = \left\{ \mathbf{I}(y) : \mathbf{I}(y) = \bigwedge_{h_{\mathbf{I}}} [\mathbf{I}(z)]\,, z \in \mathrm{SE}_x \right\}, \quad \text{erosion}$$

$$\delta_{\mathrm{SE},h_{\mathbf{I}}}\mathbf{I}(x) = \left\{ \mathbf{I}(y) : \mathbf{I}(y) = \bigvee_{h_{\mathbf{I}}} [\mathbf{I}(z)]\,, z \in \mathrm{SE}t_x \right\}, \quad \text{dilation}$$

Opening and closing

$$\gamma_{\mathrm{SE},h_{\mathbf{I}}}(\mathbf{I}) = \delta_{\mathrm{SE},h_{\mathbf{I}}}(\varepsilon_{\mathrm{SE},h_{\mathbf{I}}}(\mathbf{I})), \phi_{\mathrm{SE},h_{\mathbf{I}}}(\mathbf{I}) = \varepsilon_{\mathrm{SE},h\mathbf{I}}(\delta_{\mathrm{SE},h\mathbf{I}}(\mathbf{I}))$$

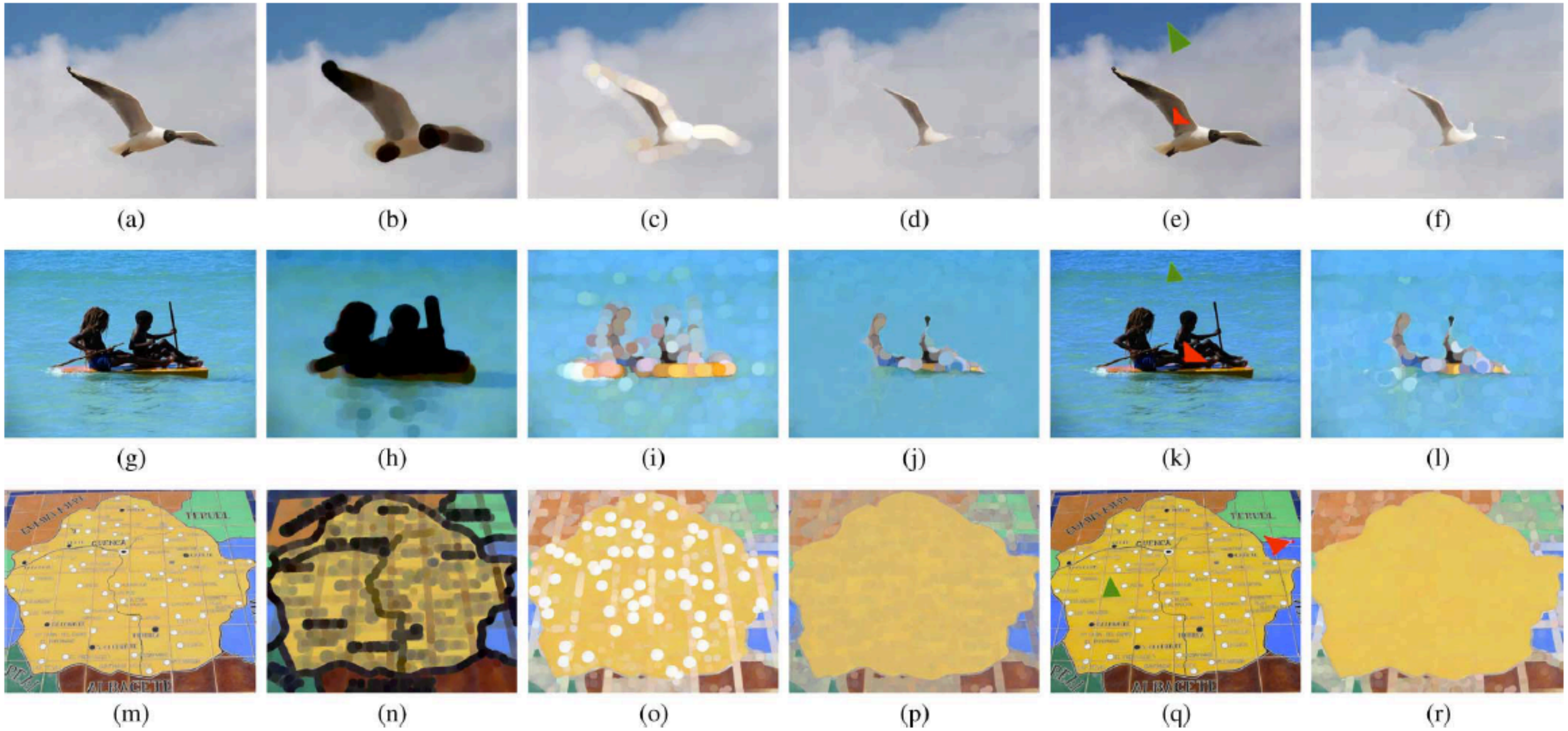


Fig. 2. Erosions by a disk of size 10 in the family of orders proposed by Barnet [6] and recent approaches from [9] and [30]. C-ordering uses the priority red $>$ green $>$ blue. Proposed P-ordering is illustrated in (e)–(k)–(q) with $k = 1000$ random projections. Supervised ordering from [30] is calculated by SVM with background/foreground sets given by green/red triangles in (e)(k)(q) respectively. Erosion in the ordering induced by the proposed P-ordering follows the physical meaning of the transformation, i.e., diminution in the size of the objects is produced. The ordering does not require a training set as supervised ordering (f)–(l)–(r). However, this intrinsic ordering is based on dichotomy background and foreground (See text for more details). (a) Original; (b) M-ordering; (c) C-ordering [9]; (d) P-ordering; (e) Training set; (f) supervised ordering [30]; (g) original; (h) M-ordering; (i) C-ordering [9]; (j) P-ordering; (k) Training set; (l) supervised ordering [30]; (m) original; (n) M-ordering; (o) C-ordering [9]; (p) P-ordering; (q) Training set; (r) supervised ordering [30].

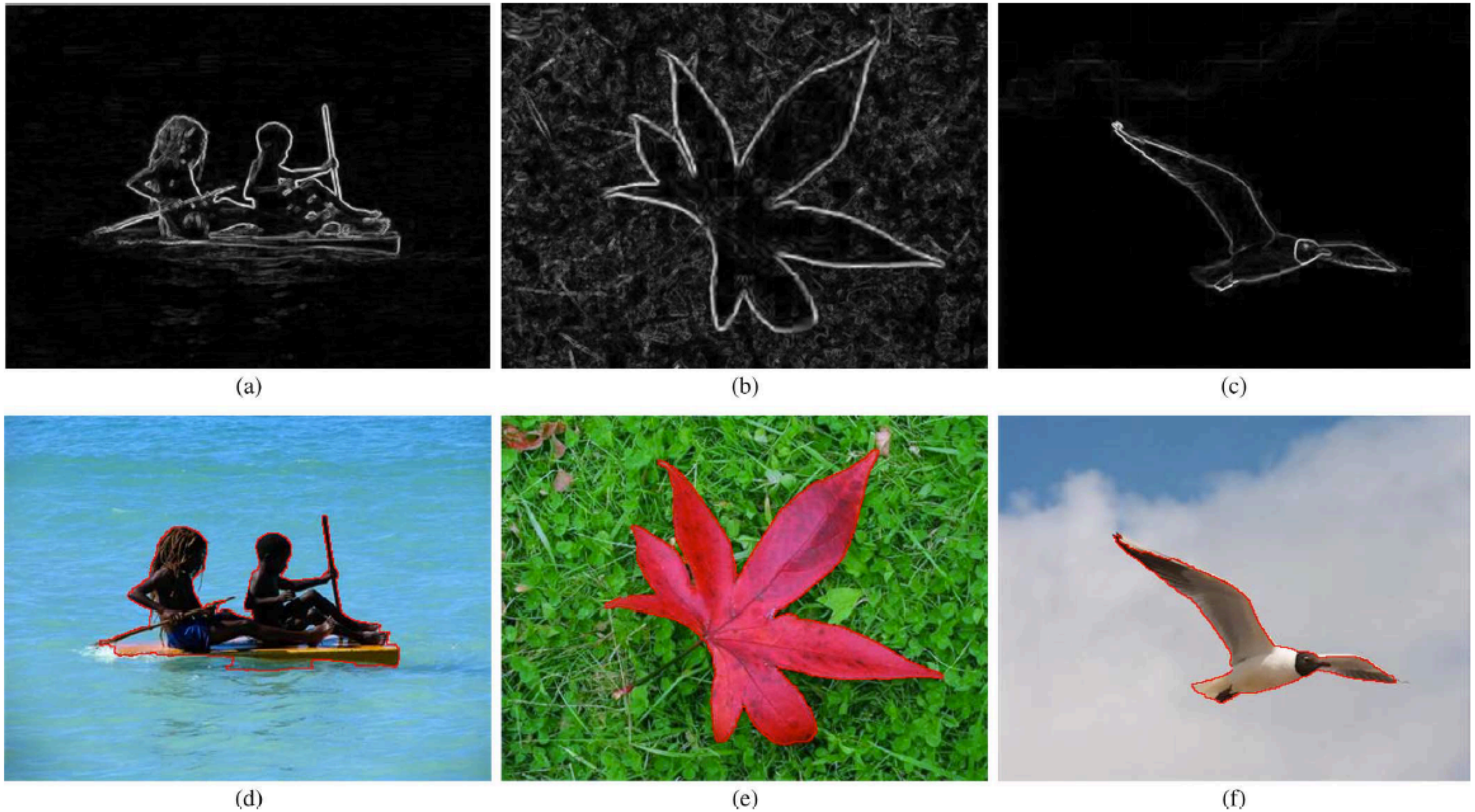


Fig. 6. h-depth gradient and segmentation by using watershed transformation (in red), where markers are calculated by selecting the minima of strong dynamics in h-depth gradient, with $t = .5$. (a) $\mathbf{\Delta}_h(\mathbf{I})$; (b) $\mathbf{\Delta}_h(\mathbf{I})$; (c) $\mathbf{\Delta}_h(\mathbf{I})$; (d) $\mathtt{WS}(\mathbf{I}, t)$; (e) $\mathtt{WS}(\mathbf{I}, t)$; (f) $\mathtt{WS}(\mathbf{I}, t)$.

Contents

- Introductory ideas and history
- Filtering
  – Fuzzy Mathematical Morphology
  – Multivariate Mathematical Morphology
- Classification
  – Fuzzy ART
  – Max-min classifiers
  – Fuzzy Lattice Neurocomputing
- Associative Morphological Memories
- Conclusions and the future

Fuzzy ART

Carpenter, Grossberg

Starting point

- It is an extensión of binary input Adaptive Resonance Theory (ART) to continuous variables in [0,1]:
  - Logical AND, intersection --> inf operator
- Coding:
  - appending the complementary $(1\text{-}x_i)$ to each input variable x_i.
- Category == Cluster

ART 1 (BINARY)	FUZZY ART (ANALOG)
CATEGORY CHOICE	
$T_j = \frac{\lvert \mathbf{I} \cap \mathbf{w}_j \rvert}{\alpha + \lvert \mathbf{w}_j \rvert}$	$T_j = \frac{\lvert \mathbf{I} \wedge \mathbf{w}_j \rvert}{\alpha + \lvert \mathbf{w}_j \rvert}$
MATCH CRITERION	
$\frac{\lvert \mathbf{I} \cap \mathbf{w} \rvert}{\lvert \mathbf{I} \rvert} \geq \rho$	$\frac{\lvert \mathbf{I} \wedge \mathbf{w} \rvert}{\lvert \mathbf{I} \rvert} \geq \rho$
FAST LEARNING	
$\mathbf{w}_j^{(new)} = \mathbf{I} \cap \mathbf{w}_j^{(old)}$	$\mathbf{w}_j^{(new)} = \mathbf{I} \wedge \mathbf{w}_j^{(old)}$
$\cap$ = logical AND intersection	$\wedge$ = fuzzy AND minimum

Fig. 2. Comparison of ART 1 and fuzzy ART.

Algorithm Elements

- Category selection based on T_j
  - It is a measure of inclusion of the input in the category

$$T_J = \max\{T_j : j = 1 \cdots N\}.$$

$$(\boldsymbol{p} \wedge \boldsymbol{q})_i \equiv \min(p_i, q_i)$$

$$T_j(\boldsymbol{I}) = \frac{|\boldsymbol{I} \wedge \boldsymbol{w}_j|}{\alpha + |\boldsymbol{w}_j|},$$

$$|\boldsymbol{p}| \equiv \sum_{i=1}^{M} |p_i|$$

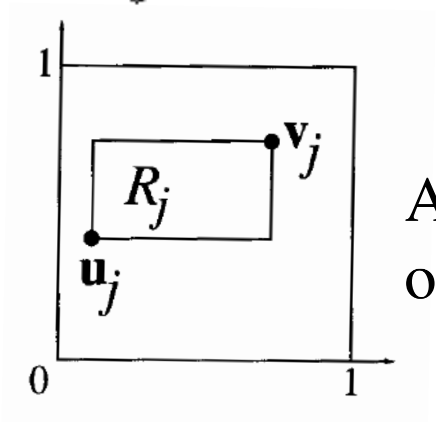
- Resonance: Vigilance parameter ρ
  - Decision about the creation of a new category
  - Measure of category compactness: inclusion of the weight $\boldsymbol{w}_J$ in the input $\boldsymbol{I}$

$$\frac{|\boldsymbol{I} \wedge \boldsymbol{w}_J|}{|\boldsymbol{I}|} \geq \rho;$$

Input accepted in the winning category

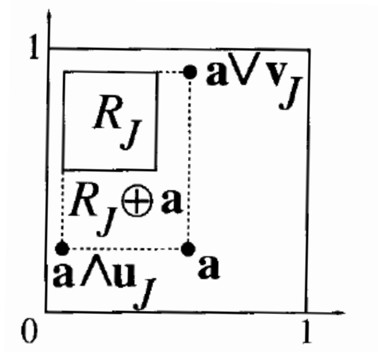
- Learning
  - Enlarging the category enclosing the new data

$$\boldsymbol{w}_J^{(\text{new})} = \beta\left(\boldsymbol{I} \wedge \boldsymbol{w}_J^{(\text{old})}\right) + (1-\beta)\boldsymbol{w}_J^{(\text{old})}.$$



(a)

After presentation of **a** (β=1)- - ->



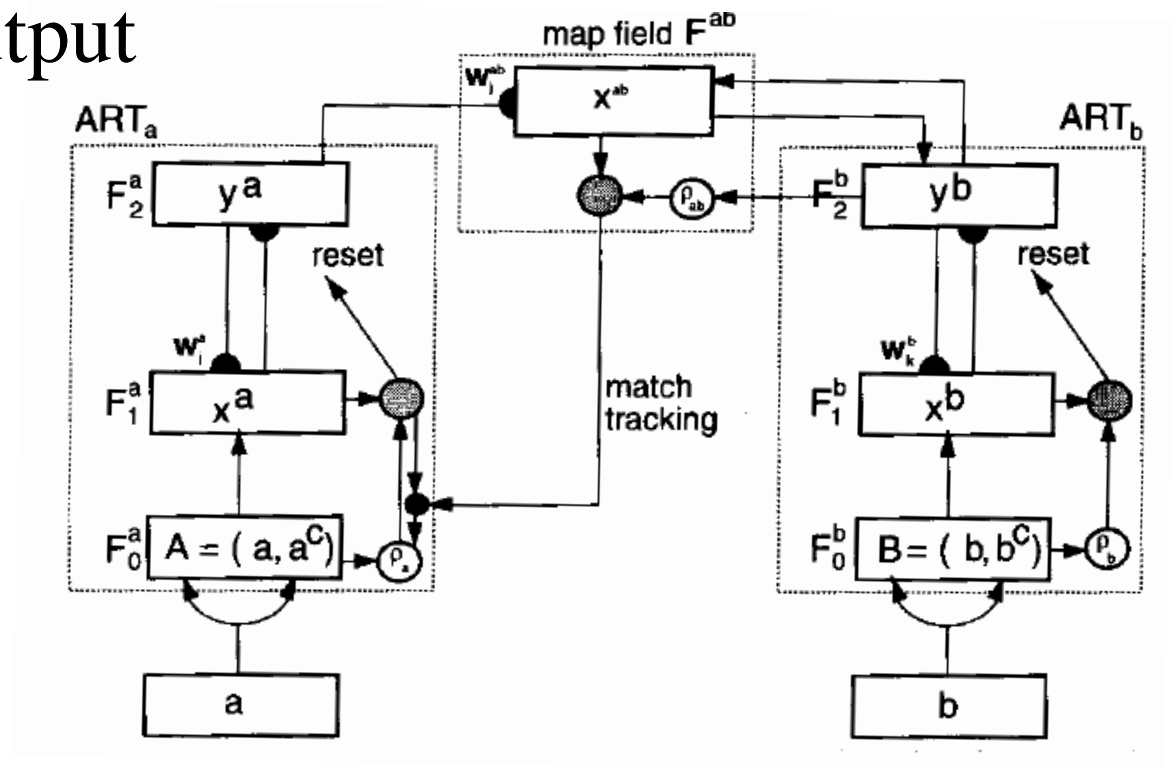
(b)

Fuzzy-ART properties

- Forms hyper-rectangular categories covering the data
- Hyper-rectangles grow monotonically in all dimensions during training
- The size of a category equals $\left|R_j\right| = M - \left|\mathbf{w}_j\right|$
- It is bounded by $\left|R_j\right| < M(1-\rho)$
- If $0 \leq \rho < 1$ the number of categories is bounded (but most times grows big!)

Supervised learning ARTMAP

- Encodes and categorizes both input and output



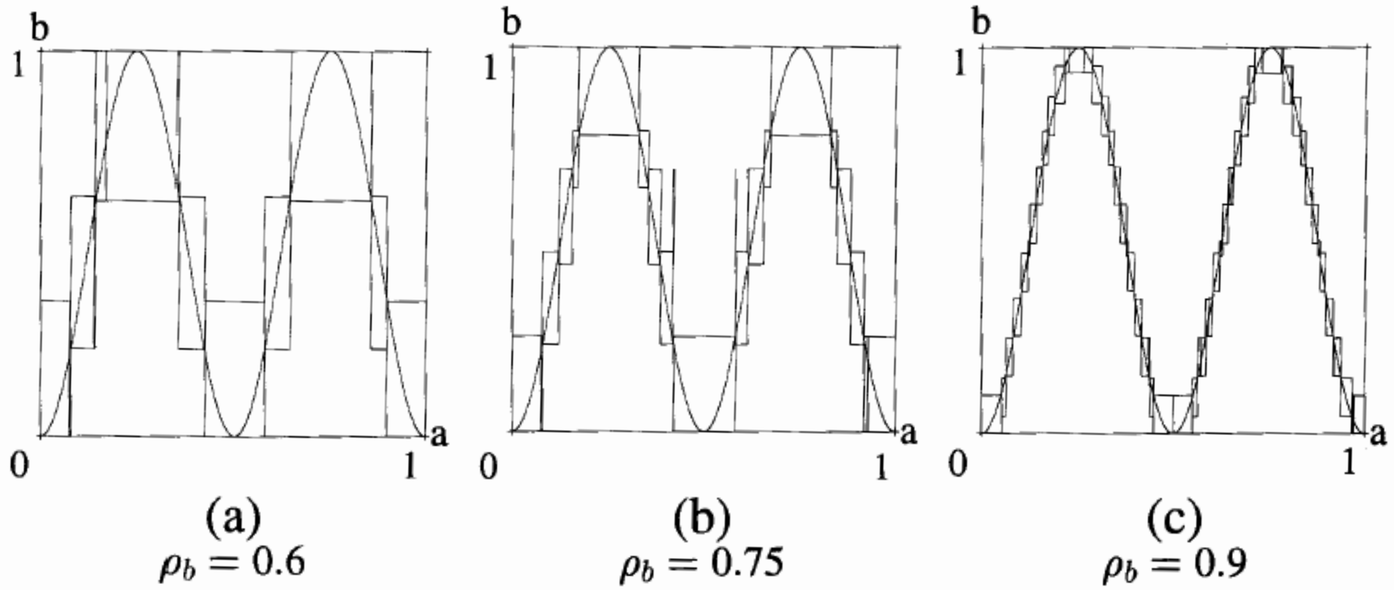


Fig. 12. Incremental approximation of a sinusoidal function for ART_b vigilance parameters, with ρ_b equal to (a) 0.6, (b) 0.75, and (c) 0.9. In each simulation the fuzzy ARTMAP system was trained on 1000 randomly chosen points $a \in [0, 1]$. Each graph shows the test set confidence intervals R_K^b selected by the test set points. The maximum lengths of these intervals are (a) 0.4, (b) 0.25, and (c) 0.1. Graph (c), with $\rho_b = 0.9$, is close to the asymptotic state of the three graphs in Fig. 11. See Table III.

Fuzzy-ARTMAP applications

- Control
- Classification and pattern recognition
- Data mining

Yang, Maragos 1995

Min-Max classifiers

Starting point

- Boolean functions in DNF

$$B(\vec{b}), \vec{b} = (b_1, \ldots, b_d) \in \{0, 1\}^d \quad b_i \in \{0, 1\}$$

- Min-max functions are obtained replacing Boolean literals by real-valued variables

$$f: [0, 1]^d \to [0, 1] \qquad x_i \in [0, 1]$$

$$f(x_1, x_2, \ldots, x_d) = \bigvee_j \bigwedge_{i \in I_j} l_i, \quad l_i \in \{x_i, 1 - x_i\}$$

- For classification a thresholding step is added

$$\theta \in [0, 1]$$

$$f_\theta(\vec{x}) = P[f(\vec{x}) \geq \theta] = \begin{cases} 1 & \text{if } f(\vec{x}) \geq \theta, \\ 0 & \text{otherwise.} \end{cases}$$

Learning

- Minimization of the Mean Square Error (MSE)

$$\mathscr{E}(t) = E[(z(t) - d(t))^2].$$

- Gradient descent on the function parameters

$$\vec{p}(t+1) = \vec{p}(t) - \mu \nabla_{\vec{p}} \mathscr{E}(t).$$

- Instantaneous error

$$\vec{p}(t+1) = \vec{p}(t) - 2\mu(z(t) - d(t))\mu \nabla_{\vec{p}} z(t)$$

- Trick
  - Assume no input variable is complemented
  - Extend the input space to 2d including the complements … Fuzzy-ART?
- Problems
  - Define parameters to allow differentiability
  - Approximate gradient of min, max, threshold

Functional form

$$h_j = \bigwedge_{i \in I_j} x_i, \quad j = 1, 2, \ldots, k$$

clause

$$y = \bigvee_{j=1}^{k} h_j$$

expression

$$z = \begin{cases} 1 & y \geq \theta, \\ 0 & y < \theta. \end{cases}$$

Decision through threshold

- How to model continuosly the conjunctive expression structure: I_j ?
  - Continuous variables m_{ij} such that
    - x_i is included in I_j if $m_{ij} \geq 0$,
    - x_i is excluded from I_j if $m_{ij} < 0$.
  - The parameters to be learnt

$$\vec{p}(t) = (\theta(t), m_{11}(t), \ldots, m_{d1}(t), \ldots, m_{dk}(t)).$$

- Derivative with respect to the threshold

$$\frac{\partial z}{\partial \theta} = \begin{cases} -\dfrac{1}{2\beta} & \text{if } |y-\theta| \le \beta, \\ 0 & \text{otherwise.} \end{cases}$$

- Where β is the width of a pulse approximating the derivative of the step function

- Derivative with respect to the structure parameters

$$\frac{\partial z}{\partial m_{ij}} = \frac{\partial z}{\partial y}\frac{\partial y}{\partial h_j}\frac{\partial h_j}{\partial m_{ij}}$$

- Implies the derivative of maximum and minimum functions.

Derivative of maximum

- Implicit formulation of maximum

$$G(y, h_1, \ldots, h_k) = \sum_{j=1}^{k} \{U_3(y - h_j) - 1\} + \frac{G_e}{2} = 0$$

$$U_3(x) = \begin{cases} 1 & \text{if } x > 0, \\ \frac{1}{2} & \text{if } x = 0, \\ 0 & \text{if } x < 0. \end{cases}$$

- Leads to the following expression

$$\frac{\partial y}{\partial h_j} \approx \begin{cases} \dfrac{1}{N_{max}} & \text{if} \quad 0 \le y - h_j \le \beta \\ 0 & \text{otherwise.} \end{cases}$$

$$N_{max} \stackrel{\Delta}{=} \text{number of } h_j\text{'s such that } y - h_j \le \beta$$
$$= \sum_{j=1}^{k} U_2(\beta - (y - h_j)).$$

Results on handwritten digit recognition

Table 1. Results for 0–1 classification problem employing *both* shape-size histograms and Fourier descriptors

Distinguishing 0's and 1's
Normalized radial size histograms and Fourier descriptors

No. of minima	Min–max % error (train)	% error (test)	Network	Neural network % error (train)	% error (test)
1	0.083	0.25	1, 1	0.083	0
3	0.083	0.25	3, 1	0.083	0
5	0.1	0.25	5, 1	0.083	0
7	0.083	0.25	7, 1	0.083	0
Normalized shape-size histograms with 2 × 2 square and Fourier descriptors					
1	3.867	2.6	1, 1	0.633	1.2
3	1.9	2.8	3, 1	0.633	0.85
5	1.083	3	5, 1	0.567	0.8
7	1.733	3	7, 1	0.533	0.55

The top two tables are generated using normalized radial histograms and Fourier descriptors, while the lower two using normalized shape-size histogram with 2 × 2 square and Fourier descriptors.

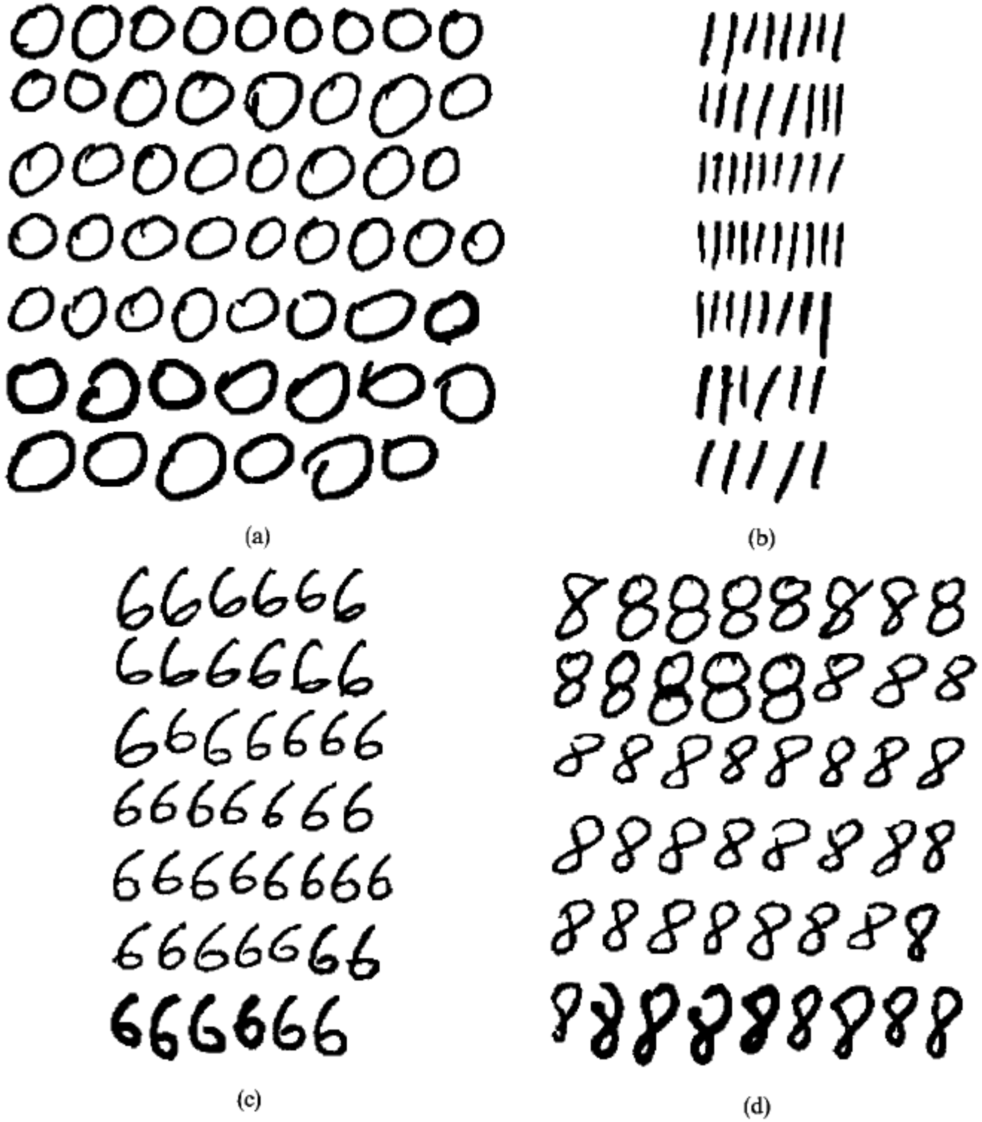


Fig. 4. Sample data from the handwritten database. (a) A collection of 0's. (b) A collection of 1's. (c) A collection of 6's. (d) A collection of 8's.

Modelling and Knowledge representation based in Lattice Theory

V. G. Kaburlasos

Starting point

- Generalizes the Fuzzy-ART and Fuzzy-ARTMAP architectures
- The Fuzzy Lattice Neurocomputing
  - Proposes an abstract representation (FIN) based on generalized interval (GI).
  - Is defined based on inclusion measures and distances on the FINs

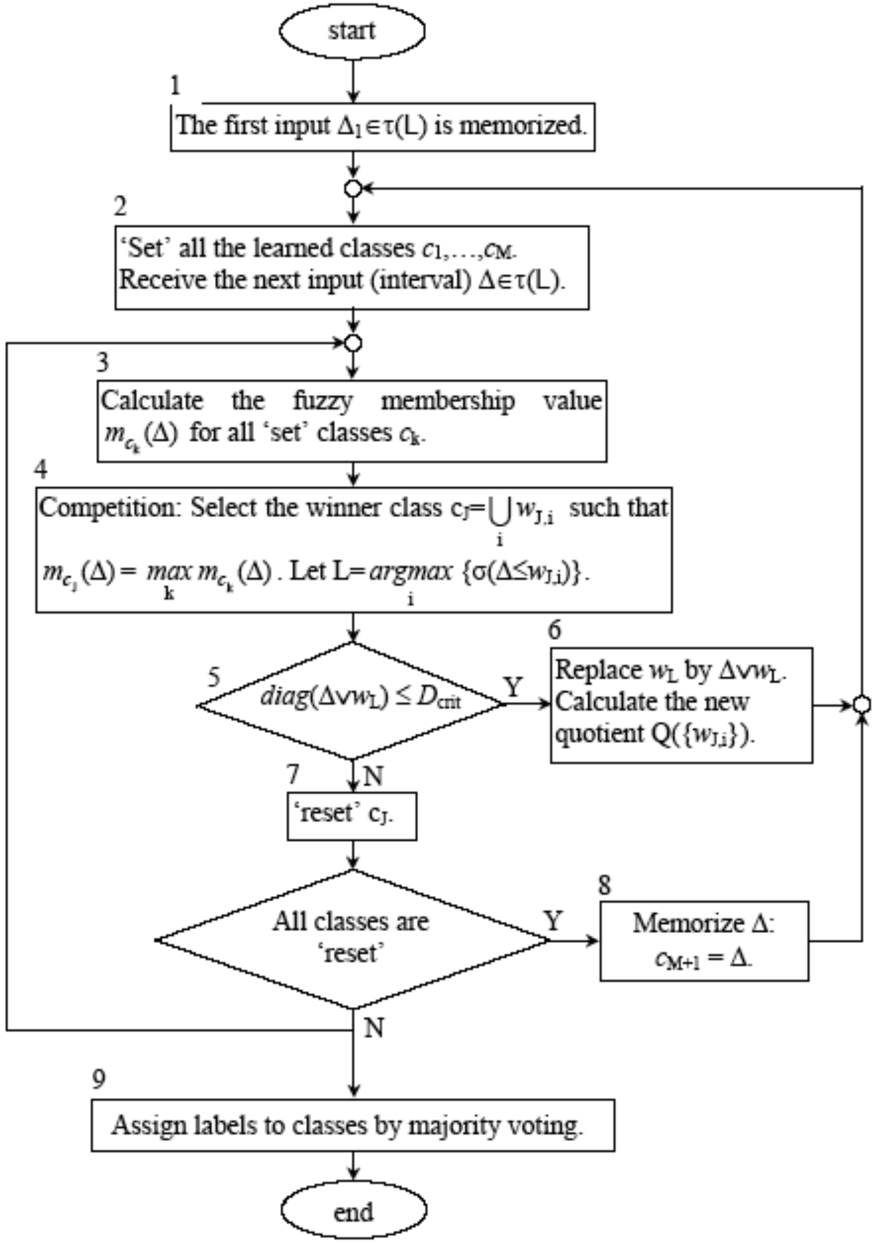


Fig.7-1 Flowchart of algorithm σ-FLN for learning (training).

Inclusion measure

$\sigma(\Delta \le w_{J,i})$

Vigilance parameter

D_{crit}.

learning

$\Delta \vee w_L$.

Advantages of σ-FLN

- Deals with data uncertainty
- Different positive valuation functions
- Deals with disparate (lattice) data types
- *Missing* and *don't care* values are treated naturally: least and greatest lattice elements.
- Learning in one step, presentation order dependent

Intervals in the unit hypercube

- Lattice interval corresponds to a hyperbox

$$\Delta = [a,b] = [(a_1,\ldots,a_N),(b_1,\ldots,b_N)] = [a_1,b_1,\ldots,a_N,b_N],$$

- Positive valuation function

$$v(w) = v(\theta(p)) + \mathrm{v}(q) = \mathrm{N} + \sum_{i=1}^{N}(q_i - p_i)$$

- Lattice join

$$\Delta \vee w = [a_1,b_1,\ldots,a_N,b_N] \vee [p_1,q_1,\ldots,p_N,q_N] =$$

$$[a_1 \wedge p_1,\ b_1 \wedge q_1,\ldots,a_N \vee p_N,\ b_N \vee q_N].$$

- Degree of inclusion

$$\sigma(\Delta \le w) = \frac{v(\theta(p)) + v(q)}{v(\theta(a \vee p) + v(b \vee q)} = \frac{N + \sum_{i=1}^{N} (v_i(q_i) - v_i(p_i))}{N + \sum_{i=1}^{N} [v_i(b_i \vee q_i) - v_i(a_i \wedge p_i)]}.$$

Algorithm 1 flrART Clustering

Assume a set $C \subset 2^{\mathfrak{I}_1^N}$; $K = |C|$; a user-defined vigilance parameter $\rho \in [0, 1]$;
for $i = 1$ to $i = n$ **do**
  Consider the next input datum $\mathbf{X}_i \in \mathfrak{I}_1^N$;
  $S \doteq C$;
  $J \doteq \underset{\substack{j \in \{1,\dots,|S|\} \\ \mathbf{W}_j \in S}}{\texttt{argmax}} \{\sigma(\mathbf{X}_i \subseteq \mathbf{W}_j)\}$;
  while $(S \neq \{\})$.and.$(\sigma(\mathbf{W}_J \subseteq \mathbf{X}_i) < \rho)$ **do**
    $S \doteq S \setminus \{\mathbf{W}_J\}$;
    $J = \underset{\substack{j \in \{1,\dots,|S|\} \\ \mathbf{W}_j \in S}}{\texttt{argmax}} \{\sigma(\mathbf{X}_i \subseteq \mathbf{W}_j)\}$;
  end while
  if $S = \{\}$ **then**
    $C \doteq C \cup \{\mathbf{X}_i\}$;
    $K \doteq K + 1$;
  else
    $\mathbf{W}_J \doteq \mathbf{W}_J \dot{\cup} \mathbf{X}_i$;
  end if
end for

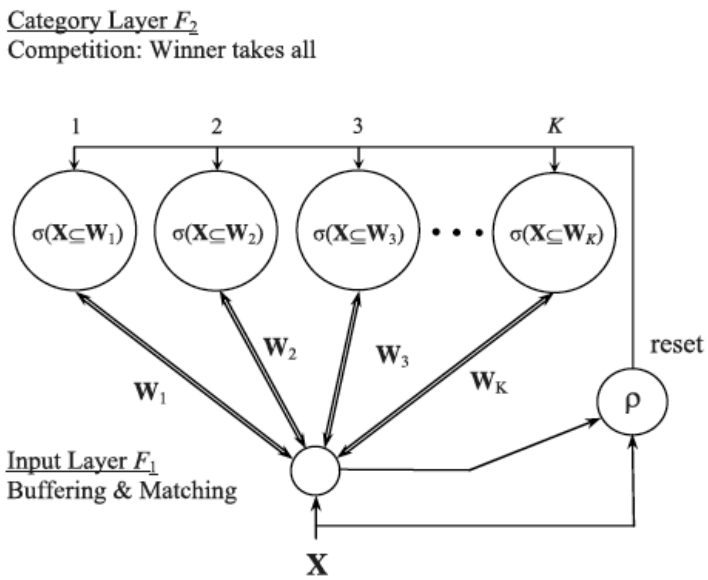


Fig. 2. flrART neural architecture for clustering, where an input pattern $\mathbf{X}$ is in the lattice $(\mathfrak{I}_1^N, \subseteq)$ of intervals.

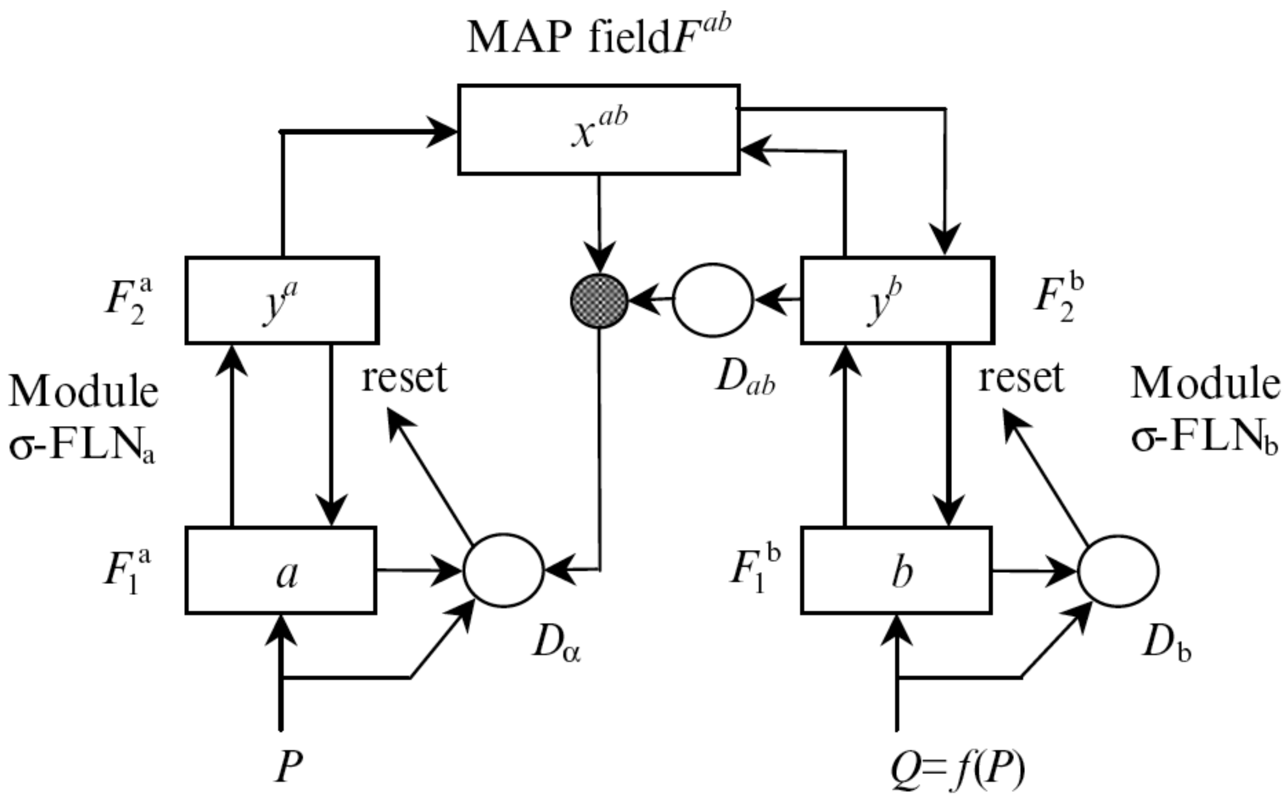


Fig.7-4 The σ-FLNMAP neural network for inducing a function f: τ(L)→τ(K), where both L and K are mathematical lattices.

Generalization

- Positive Valuation function on a lattice (L,≤) satisfies

$$v(x)+v(y)=v(x\wedge y)+v(x\vee y)$$

$$x<y\Rightarrow v(x)<v(y)$$

- A positive valuation in a lattice (L,≤) induces a metric (distance) $d:L\times L\rightarrow R_0^+$

$$d(x,y)=v(x\vee y)-v(x\wedge y)$$

- An inclusion measure is a function $\sigma : L \times L \to [0,1]$ satisfying

$$\text{(IM1)} \quad \sigma(x,x) = 1, \forall x \in L$$

$$\text{(IM2)} \quad x \wedge y < x \Rightarrow \sigma(x,y) < 1$$

$$\text{(IM3)} \quad u \leq w \Rightarrow \sigma(x,u) \leq \sigma(x,w)$$

- If v is a positive valuation in lattice (L,≤) then both expressions are inclusion measures

$$\text{(a)} \quad k(x,u) = \frac{v(u)}{v(x \vee u)} \qquad \text{(b)} \quad k(x,u) = \frac{v(x \wedge u)}{v(x)}$$

Fuzzy Interval Numbers (FIN)

- A **FIN** is a function $F:(0,1]\rightarrow M$ such that

  (1) $F(h)\in M^h$

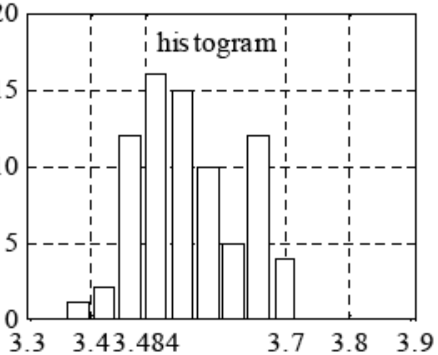
  (2) either $F(h)\in M_+^h$ or $F(h)\in M_-^h$

  (3) $h_1 \leq h_2 \Rightarrow \{x:F(h_1)\neq 0\}\supseteq\{x:F(h_2)\neq 0\}$

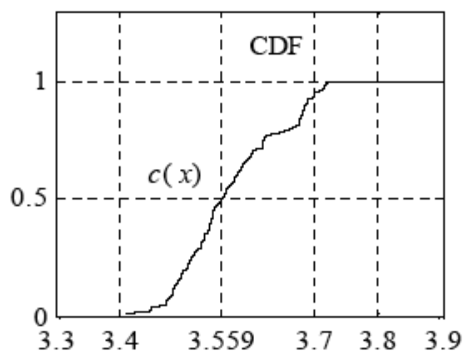
- where M^h denotes the set of generalized intervals of heigh h. It is a **lattice ordered linear space**.

- FINs can be models of
  - Real numbers
  - Intervals
  - Fuzzy numbers
  - Probability distributions
- FINs inherit valuation, inclusion, metric functions from the set of generalized intervals

Probability distribution FIN

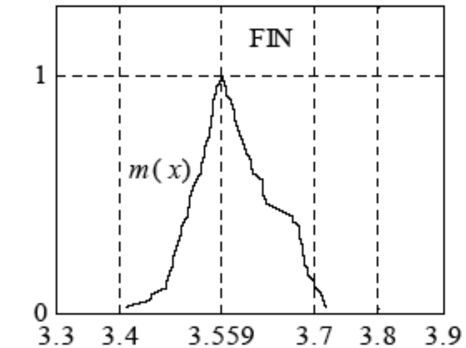


(a)



(b)

$$m(x)=\begin{cases}0.5c(x), & x\le 3.559\\ 1\text{-}0.5c(x), & x>3.559\end{cases}$$



(c)

Operations on FINs

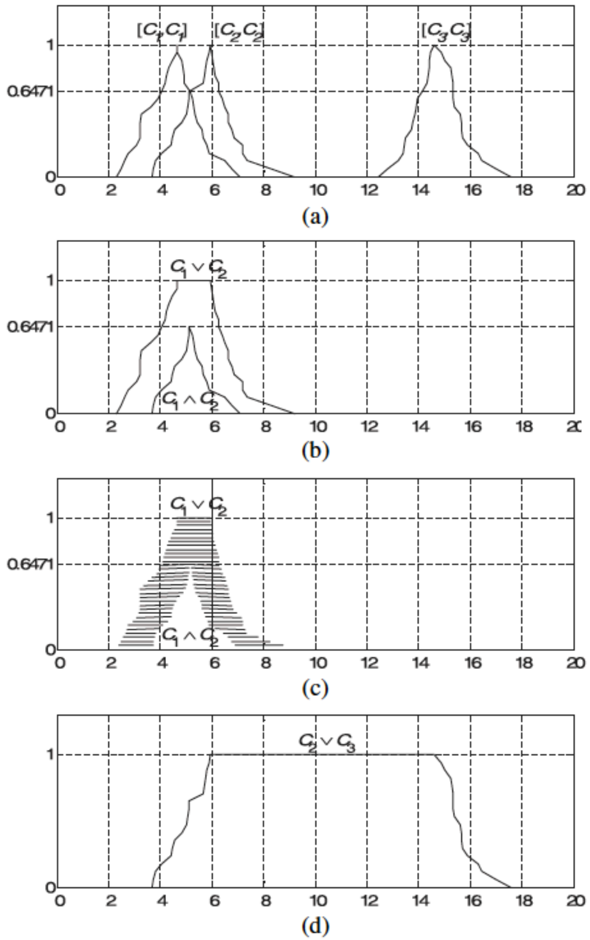


Fig. 1. Demonstrating the lattice join ($\curlyvee$) operation between trivial Type-2 INs. (a) Trivial Type-2 INs $[C_1, C_1] = \mathbb{C}_1$, $[C_2, C_2] = \mathbb{C}_2$, and $[C_3, C_3] = \mathbb{C}_3$. (b) Type-2 IN $\mathbb{C}_1 \curlyvee \mathbb{C}_2 = [C_1 \curlywedge C_2, C_1 \curlyvee C_2]$ is shown in its membership-function representation. (c) Type-2 IN $\mathbb{C}_1 \curlyvee \mathbb{C}_2 = [C_1 \curlywedge C_2, C_1 \curlyvee C_2]$ is shown again, this time in its (equivalent) interval representation for $L = 32$ different levels spaced uniformly over the interval [0, 1] on the vertical axis. (d) Type-2 IN $\mathbb{C}_2 \curlyvee \mathbb{C}_3 = [C_2 \curlywedge C_3, C_2 \curlyvee C_3] = [\emptyset, C_2 \curlyvee C_3]$.

Applications

- Classification and clustering
  - Benchmark problems
  - Epidural surgery planification
  - Orthopedics bone drilling
  - Ambient ozone estimation
  - Prediction of industrial sugar production

Lattice Computing Extension of the FAM NeuralClassifier for Human Facial Expression Recognition

Vassilis G. Kaburlasos, Stelios E. Papadakis, and George A. Papakostas

IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, VOL. 24, NO. 10, OCTOBER 2013

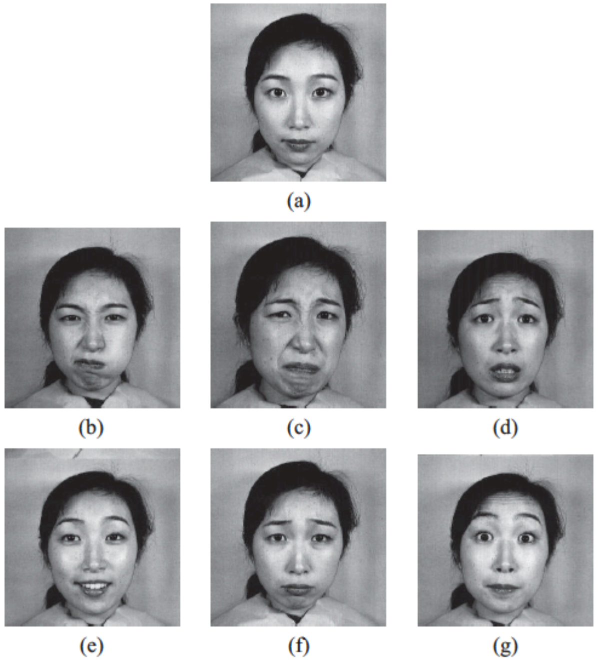


Fig. 4. Seven different facial expressions from the JAFFE benchmark dataset, including (a) neutral, (b) angry, (c) disgusted, (d) fear, (e) happy, (f) sad, and (g) surprise.

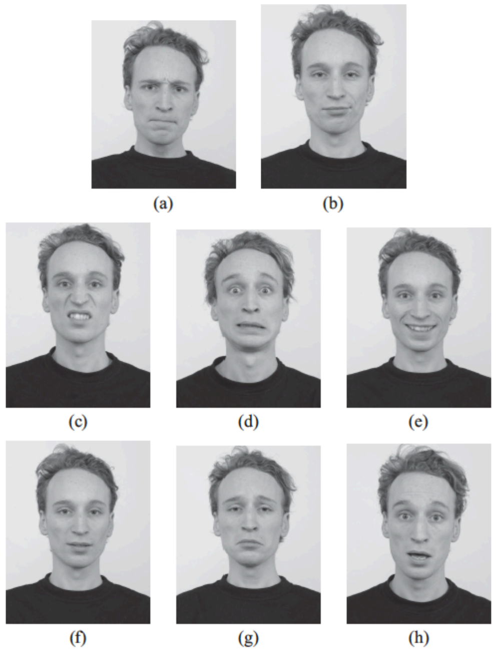


Fig. 5. Eight different emotional expressions from the RADBOUD benchmark dataset, including (a) angry, (b) contemptuous, (c) disgusted, (d) fear, (e) happy, (f) neutral, (g) sad, and (h) surprise.

- Features:
  - 16- dimensional feature vector moments
    - Zernike, Pseudo–Zernike, Fourier–Mellin, Legendre, Tchebichef, or Krawtchouk moments
  - 6x16 dimensional feature vectors with all the moment features

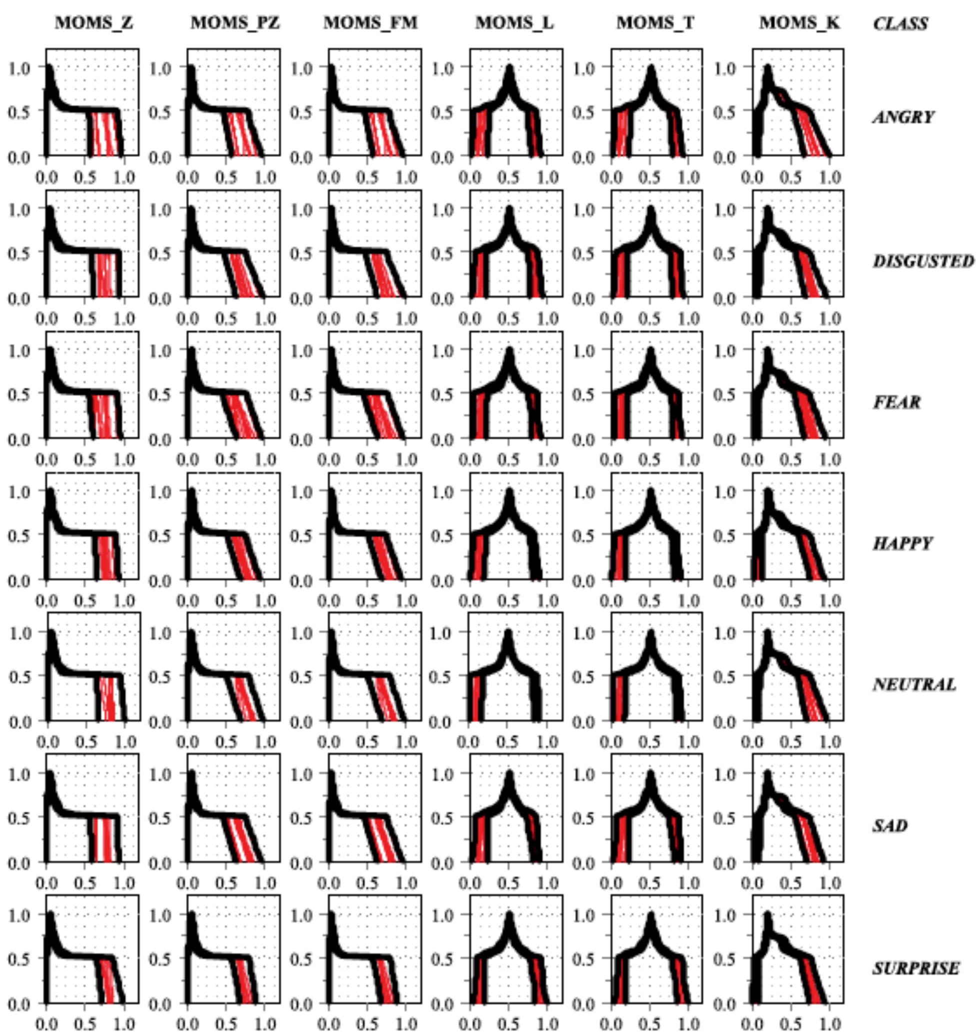


Fig. 6. A row of the 7 × 7 Table above (excluding the header) displays one 6-dimensional Type-2 IN induced for each of the seven human facial expressions (classes) of the JAFFE benchmark dataset. One Type-2 IN corresponds to one kind of moment. At the end of a row, the corresponding class name is shown.

TABLE I

GENERALIZATION RATE (%) STATISTICS REGARDING THE JAFFE TESTING DATA IN 10 COMPUTATIONAL EXPERIMENTS USING SEVERAL CLASSIFIERS AND SIX DIFFERENT KINDS OF MOMENTS, CONCATENATED

Classifier Name	Min	Max	Ave	Std
kNN ($k = 1$)	40.91	94.74	67.68	15.82
Naive Bayes	18.18	52.63	36.80	10.03
Classification tree	31.82	47.37	40.02	5.67
Neural network (50)	18.18	59.09	37.27	13.52
FAM	50.00	90.00	68.87	13.49
flrFAM	50.00	86.36	69.54	12.31

TABLE II

GENERALIZATION RATE (%) STATISTICS REGARDING THE RADBOUD TESTING DATA IN 10 COMPUTATIONAL EXPERIMENTS USING SEVERAL CLASSIFIERS AND SIX DIFFERENT KINDS OF MOMENTS, CONCATENATED

Classifier Name	Min	Max	Ave	Std
kNN ($k = 1$)	22.22	46.30	35.74	7.51
Naive Bayes	35.19	57.41	48.15	7.04
Classification tree	27.78	40.74	34.07	4.20
Neural network (50)	11.11	64.81	45.74	15.81
FAM	27.77	44.44	37.40	6.03
flrFAM	35.18	50.00	43.14	4.86

Contents

- Introductory ideas and history
- Filtering
  - Fuzzy Mathematical Morphology
  - Multivariate Mathematical Morphology
- Classification
  - Fuzzy ART
  - Max-min classifiers
  - Fuzzy Lattice Neurocomputing
- Associative Morphological Memories
- Conclusions and the future

Associative Morphological Memories

Ritter, Sussner

Starting point

- Linear neuron

$$\tau_i(t+1) = \sum_{j=1}^{n} a_j(t) \cdot w_{ij} \qquad a_i(t+1) = f(\tau_i(t+1) - \theta_i)$$

- Matrix notation

$$T(t+1) = W \cdot \mathbf{a}(t)$$

$$\mathbf{a}(t) = (a_1(t), \cdots, a_n(t))',$$

$$T(t+1) = (\tau_1(t+1), \cdots, \tau_n(t+1))',$$

- Morphological dilative neuron:

$$\tau_i(t+1) = \bigvee_{j=1}^{n} a_j(t) + w_{ij}$$

- Matrix notation: max product $T(t+1) = W \boxdot \mathbf{a}(t)$

$$C = A \boxdot B$$

$$c_{ij} = \bigvee_{k=1}^{p} a_{ik} + b_{kj} = (a_{i1} + b_{1j}) \vee (a_{i2} + b_{2j}) \vee \cdots \vee (a_{ip} + b_{pj}).$$

- Morphological erosive neuron:

$$\tau_i(t+1) = \bigwedge_{j=1}^{n} a_j(t) + w_{ij}$$

- Matrix notation: min-product $T(t+1) = W \boxtimes \mathbf{a}(t)$

$$C = A \boxtimes B$$

$$c_{ij} = \bigwedge_{k=1}^{p} a_{ik} + b_{kj} = (a_{i1} + b_{1j}) \wedge (a_{i2} + b_{2j}) \wedge \cdots \wedge (a_{ip} + b_{pj}).$$

Morphological associative memories

- Hopfield associative memory: given an input **x** recalls response **y** as

$$\mathbf{y} = W \cdot \mathbf{x}.$$

- To store *k* vector pairs

$(\mathbf{x}^1, \mathbf{y}^1), \cdots, (\mathbf{x}^k, \mathbf{y}^k)$, where $\mathbf{x}^\xi \in R^n$ and $\mathbf{y}^\xi \in R^m$

$$W = \sum_{\xi=1}^{k} \mathbf{y}^\xi \cdot (\mathbf{x}^\xi)'.$$

- The Hopfield associative memory provides **perfect recall** if the input patterns are orthogonal
- If they are not orthogonal, the recall is corrupted by crosstalk noise.

- Morphological Associative Memories
- Construction with a single pair:

$$W = \mathbf{y} \boxed{\vee} (-\mathbf{x})' ;$$

- Recall (perfect):

$$W \boxed{\vee} \mathbf{x} = \mathbf{y}$$

- Given a set of input-output patterns

$$\{(\mathbf{x}^\xi, \mathbf{y}^\xi) \quad : \xi = 1, \cdots, k\}$$

- Define: (X, Y),

$$X = (\mathbf{x}^1, \cdots, \mathbf{x}^k) \qquad Y = (\mathbf{y}^1, \cdots, \mathbf{y}^k).$$

- Two natural morphological memories

$$W_{XY} = \bigwedge_{\xi=1}^{k} [\mathbf{y}^\xi \times (-\mathbf{x}^\xi)'] \quad \text{and} \quad M_{XY} = \bigvee_{\xi=1}^{k} [\mathbf{y}^\xi \times (-\mathbf{x}^\xi)'].$$

- Basic recall property:
  - the erosive and dilative memory recalls bound the exact response

$$W_{XY} \leq \mathbf{y}^{\xi} \times (-\mathbf{x}^{\xi})' \leq M_{XY}$$

$$W_{XY} \boxtimes \mathbf{x}^{\xi} \leq \mathbf{y}^{\xi} \leq M_{XY} \boxtimes \mathbf{x}^{\xi}$$

$$W_{XY} \boxtimes X \leq Y \leq M_{XY} \boxtimes X.$$

- Conditions for perfect recall

Theorem 2: W_{XY} is ☑-perfect for (X, Y) if and only if for each $\xi = 1, \cdots, k$, each row of the matrix $[\mathbf{y}^{\xi} \times (-\mathbf{x}^{\xi})'] - W_{XY}$ contains a zero entry. Similarly M_{XY} is ◬-perfect for (X, Y) if and only if for each $\xi = 1, \cdots, k$, each row of the matrix $M_{XY} - [\mathbf{y}^{\xi} \times (-\mathbf{x}^{\xi})']$ contains a zero entry.

Autoassociative memories

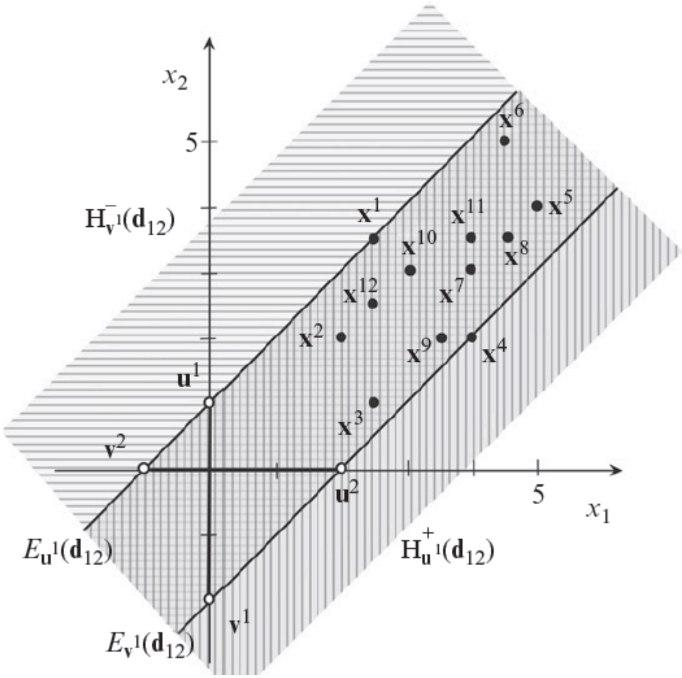
- When $X=Y$, memories W_{XX} and M_{XX} are called autoassociative.
- They have perfect recall and unlimited capacity

$$W_{XX} \boxtimes X = X \text{ and } M_{XX} \boxtimes X = X.$$

- Recalling converges in one step

Fixed points of M_{XX} and W_{XX}[a]

[a] G.X.Ritter, G.Urcid, "Lattice algebra approach to endmember determination in hyperspectral imagery," in P. Hawkes (Ed.), Advances in imaging and electron physics, Vol. 160, 113–169. Elsevier, Burlington, MA (2010)



Noise

- Memory W_{XX} is robust to erosive noise and sensitive to dilative noise
- Memory M_{XX} is robust to dilative noise and sensitive to erosive noise

$\tilde{\mathbf{x}}^{\gamma} \leq \mathbf{x}^{\gamma}$ Erosive noise

$\tilde{\mathbf{x}}^{\gamma} \geq \mathbf{x}^{\gamma}$ Dilative noise

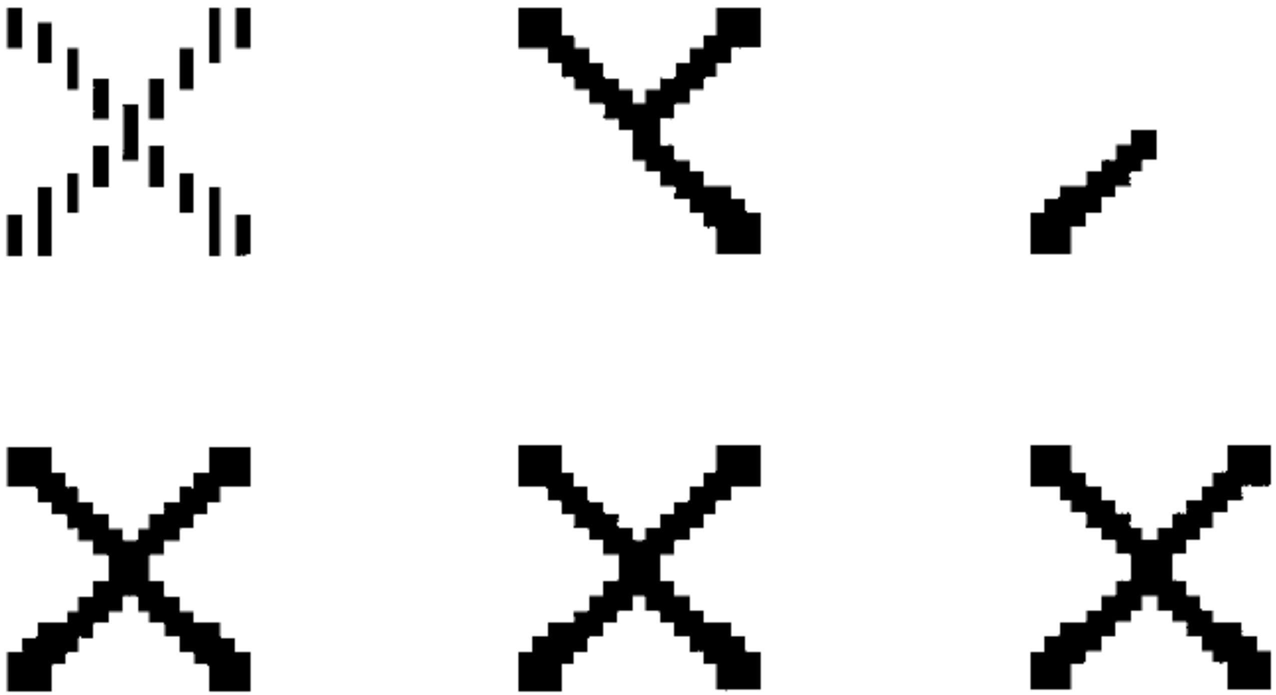


Fig. 4. The top row shows the corrupted input patterns and the bottom row the corresponding output patterns of the morphological memory W_{XX}.

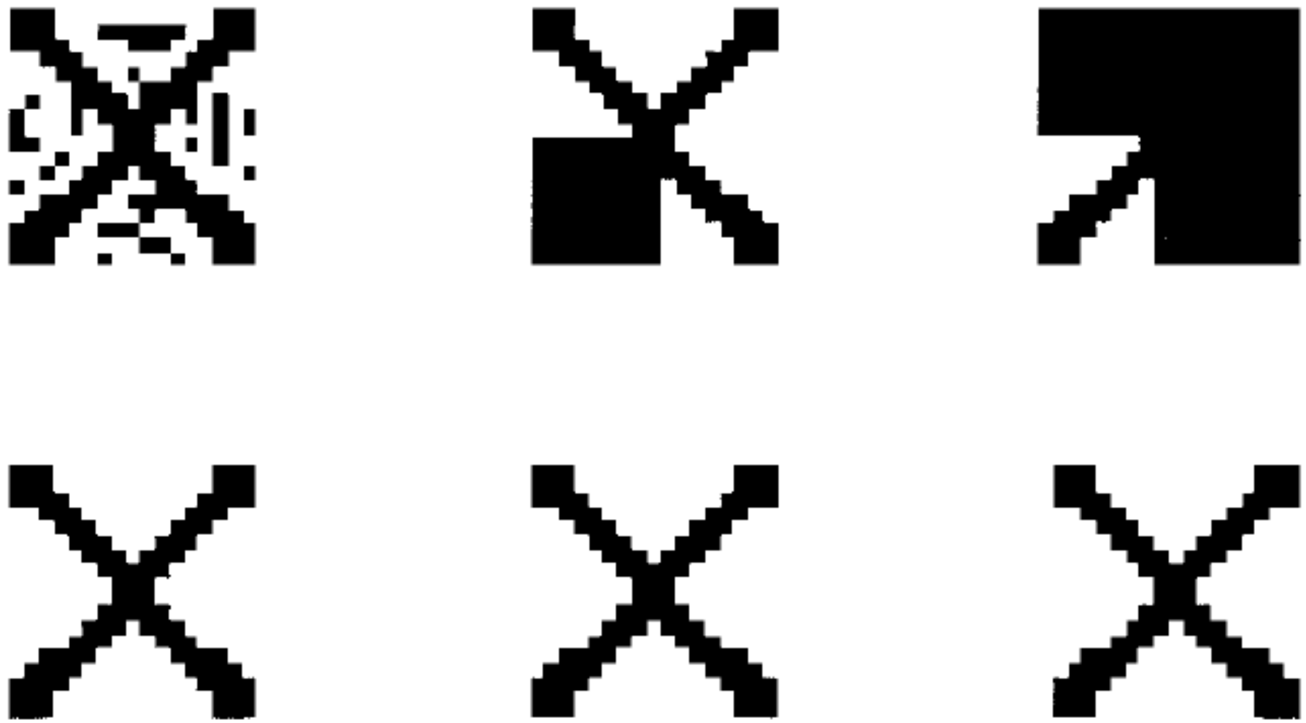


Fig. 5. The top row shows the corrupted input patterns and the bottom row the corresponding output patterns of the morphological memory M_{XX}.

Approaches to solve the noise problem

- Definition of kernels

Definition 2: Let $Z = (\mathbf{z}^1, \mathbf{z}^2, \cdots, \mathbf{z}^k)$ be an $n \times k$ matrix. We say that Z is a *kernel* for (X, Y) if and only if the following two conditions are satisfied:

1. $M_{ZZ} \boxed{\triangle} X = Z$;
2. $W_{ZY} \boxed{\triangledown} Z = Y$.

It follows that if Z is a kernel for (X, Y), then

$$W_{ZY} \boxed{\triangledown} (M_{ZZ} \boxed{\triangle} X) = W_{ZY} \boxed{\triangledown} Z = Y.$$

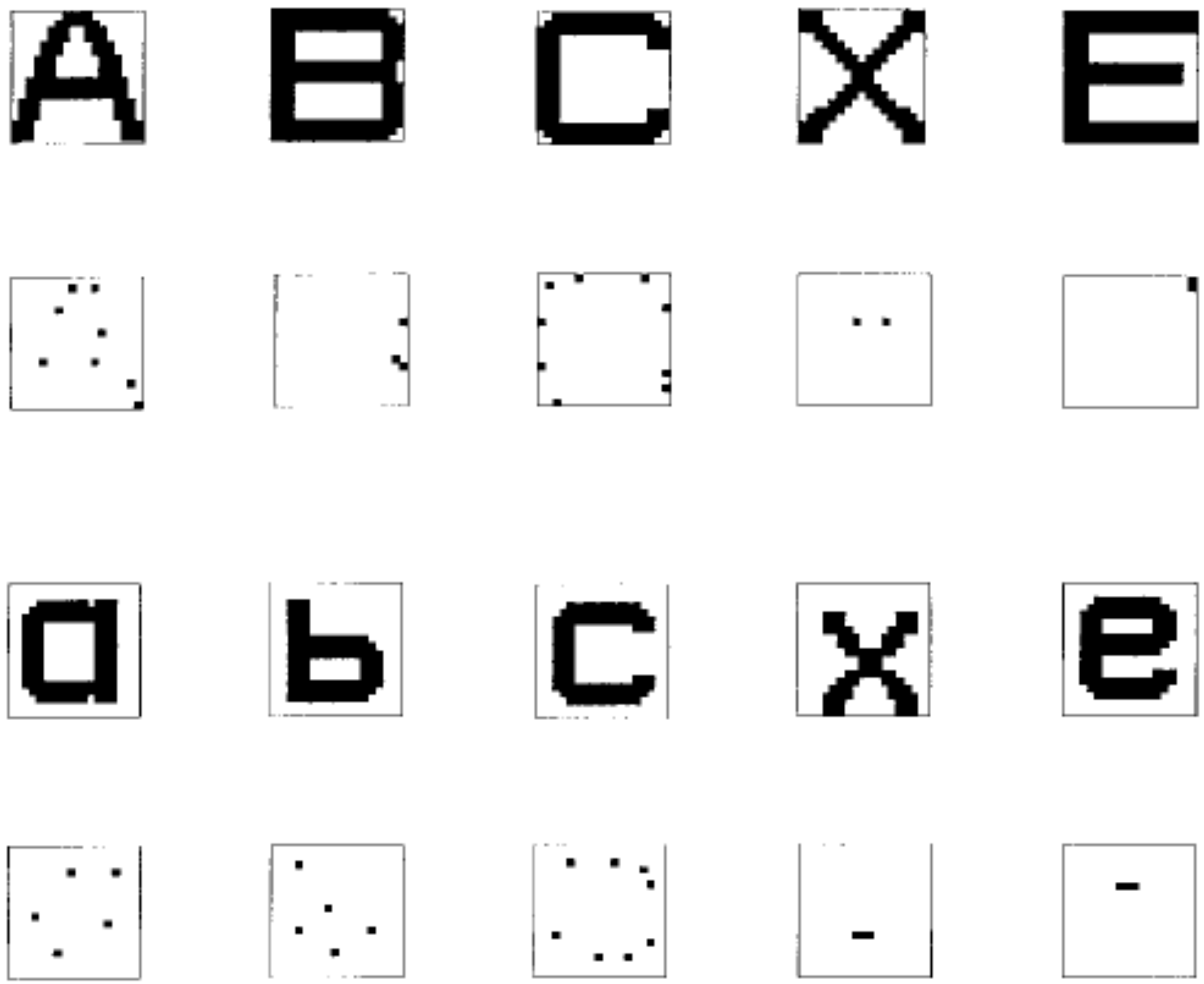


Fig. 6. An example of kernel images. The kernel image corresponding toa particular letter image is the image directly below the letter image.

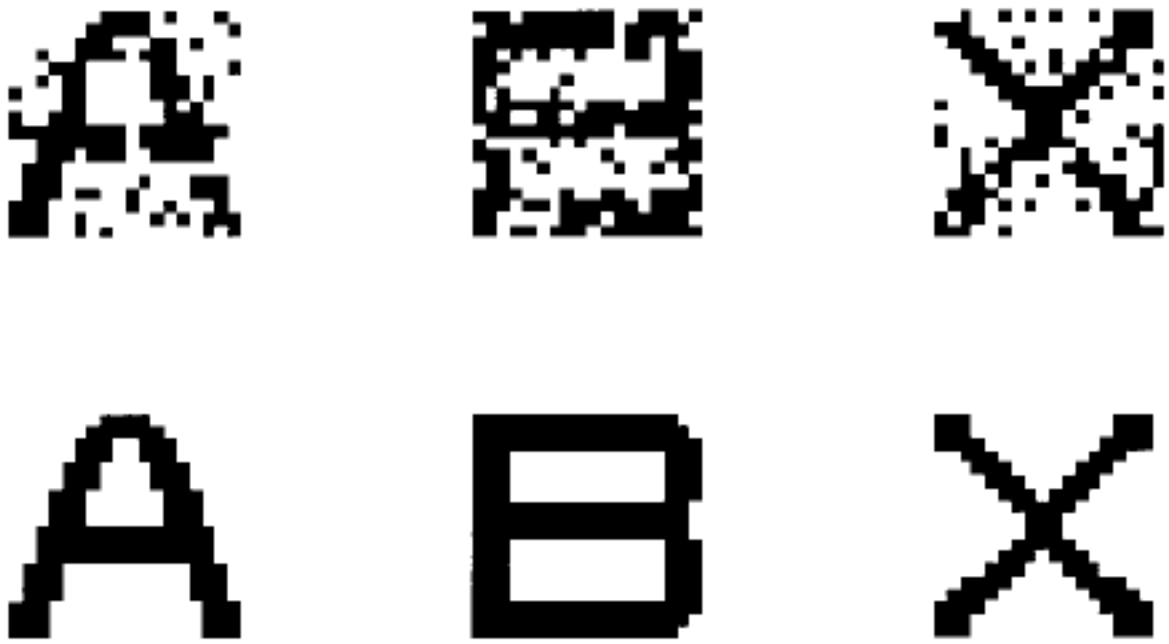
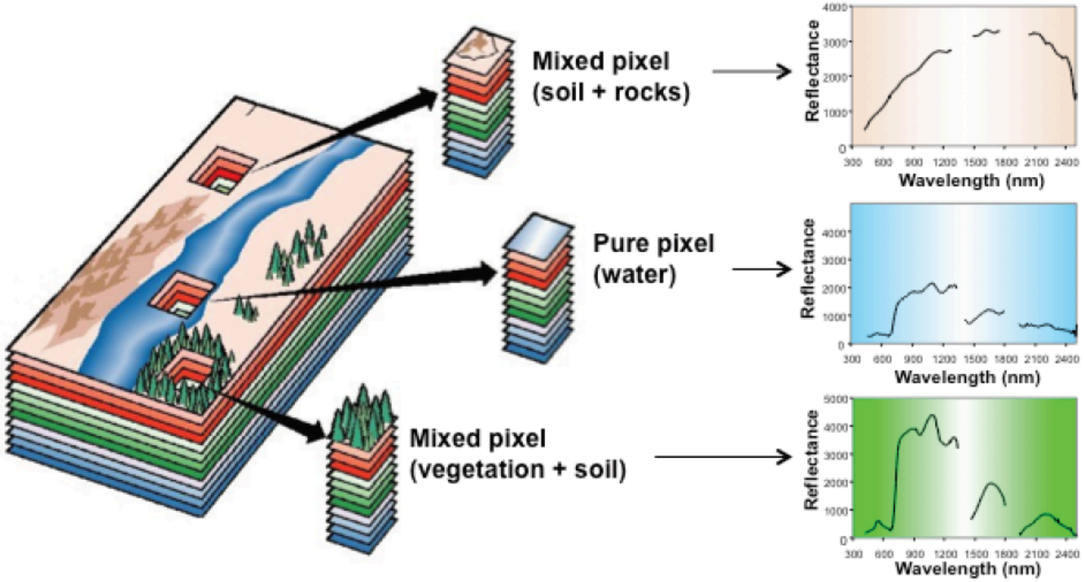


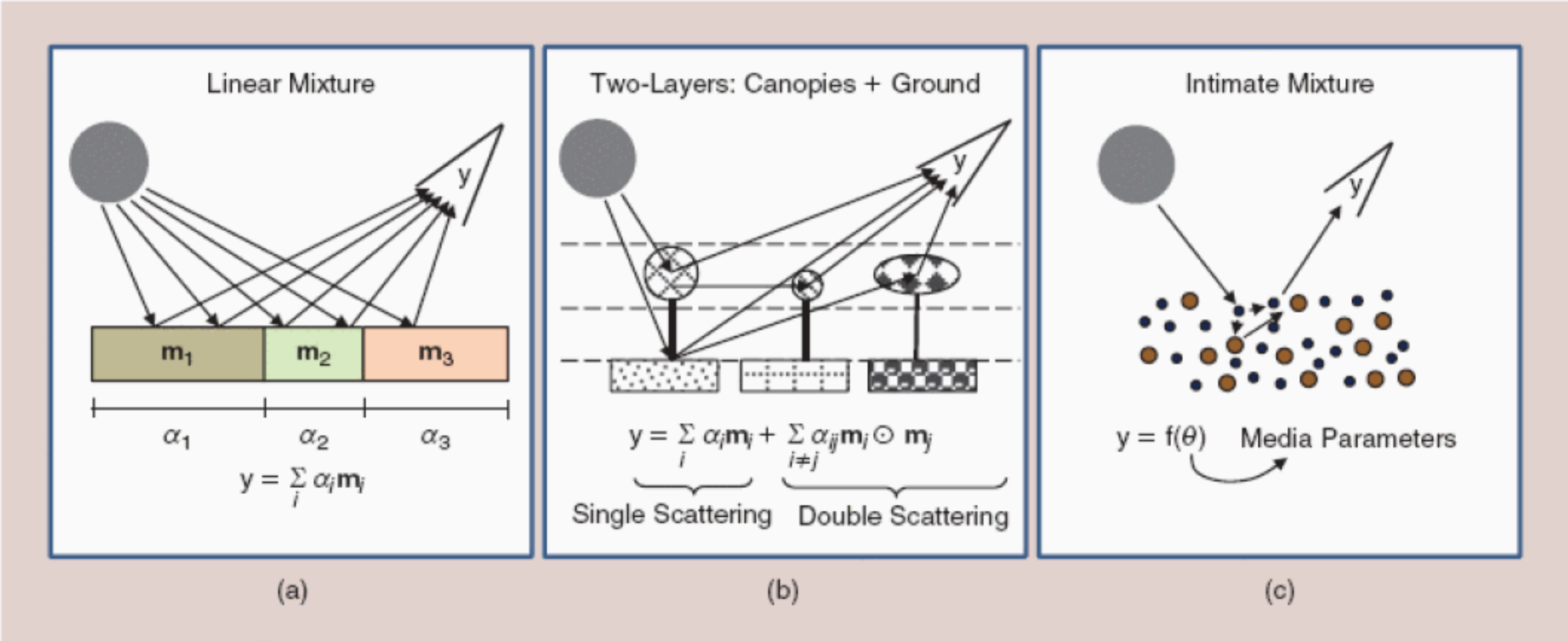
Fig. 7. An example of the behavior of the memory $\{input \rightarrow M_{ZZ} \rightarrow W_{ZY} \rightarrow output\}$. The memory was trained using the ten exemplars shown in Fig. 2. Presenting the memory with the corrupted patterns of the letters A, B, and X resulted in perfect recall (lower row). Each letter was corrupted by randomly reversing each bit with a probability of 0.15.

Application to hyperspectral images

Hyperspectral image definition



Spectral mixing

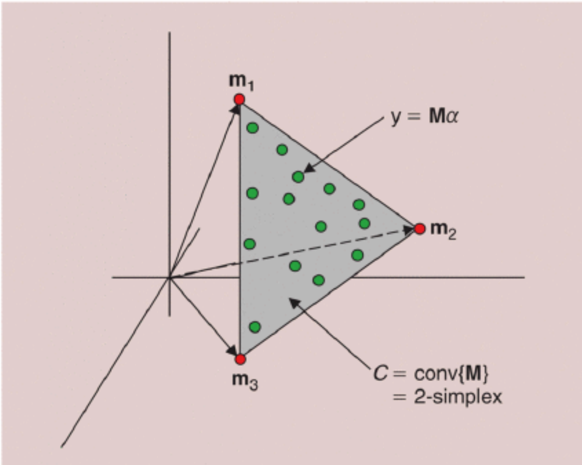


Linear mixing model

- The linear mixing model

$$x(i,j) \approx \sum_{k=1}^{p} e_k \cdot \Phi_k(i,j) + w = E \cdot \Phi + w \qquad (1)$$

with $\sum_{k=1}^{p} \Phi_k(i,j) = 1$ and $\Phi \geq 0$; E is the set of endmembers.



Linear Unmixing

The statement of the problem

$$\min_{\mathbf{M},\mathbf{A}} \|\mathbf{Y} - \mathbf{M}\mathbf{A}\|_F^2 \text{ subject to} : \mathbf{A} \geq \mathbf{0}, \mathbf{1}_p^T \mathbf{A} = \mathbf{1}_n,$$

Endmember induction Algorithm

Definition

Endmember Induction algorithms (EIA): extracting a set of endmembers E from the data X

- Types of EIA
  - Geometric: searching for simplex covering
  - Algebraic (PCA, ICA, NNMF)
  - Lattice computing: equivalence between lattice independence and affine independence

Ritter's EIA

Algorithm 2 Endmember Threshold Selection Algorithm (ETSA) based on [27,28]

(1) Given a set of vectors $X = \left\{ \mathbf{x}^1, ..., \mathbf{x}^k \right\} \subset \mathbb{R}^n$ compute the min and max auto-associative memories W_{XX} M_{XX} from the data. Their column vector sets W and M will be the candidate endmembers.

(2) Register W and M relative to the data set adding the maximum and minimum values of the data dimensions (bands in the hyperspectral image). Obtain $\overline{W}$ and $\overline{M}$ as follows:

  (a) Compute $u_i = \bigvee_{\xi=1}^{n} x_i^{\xi}$ and $v_i = \bigwedge_{\xi=1}^{n} x_i^{\xi}$.

  (b) Compute $\overline{\mathbf{m}}^i = \mathbf{m}^i + v_i$ and $\overline{\mathbf{w}}^i = \mathbf{w}^i + u_i$

(3) Remove lattice dependent vectors from the joint set $\overline{W} \bigcup \overline{M}$.

(4) Compute the standard deviation along each dimension of the candidate endmember vectors, denoted by the vector $\overrightarrow{\sigma} = \{\sigma_1, \ldots, \sigma_n\}$.

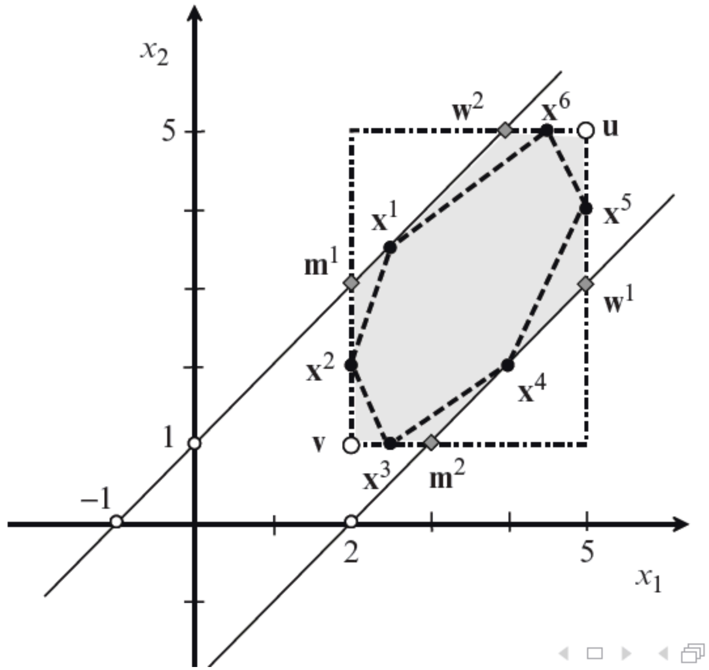
(5) Assume the first vector in the set $\mathbf{v}_1 \in \overline{W} \bigcup \overline{M}$ as the first endmember, $E = \{\mathbf{v}_1\}$

(6) Iterate for the remaining vectors $\mathbf{v} \in \overline{W} \bigcup \overline{M}$

  (a) If $\|\mathbf{v} - \mathbf{e}\| < \gamma \overrightarrow{\sigma}$ for any $\mathbf{e} \in E$ then discard $\mathbf{v}$ otherwise include $\mathbf{v}$ in E

Convex Polytope from Ritter's EIA[a]

[a] G.X.Ritter, G.Urcid, "Lattice algebra approach to endmember determination in hyperspectral imagery," in P. Hawkes (Ed.), Advances in imaging and electron physics, Vol. 160, 113–169. Elsevier, Burlington, MA (2010)



Graña's EIA[a]

[a]M. Graña, I. Villaverde, J.O. Maldonado, C. Hernandez, Two lattice computing approaches for the unsupervised segmentation of hyperspectral images, Neurocomputing 72:2111–2120 (2009)

Algorithm 3 Endmember Induction Heuristic Algorithm (EIHA)

(1) Shift the data sample to zero mean $\{\mathbf{f}^c(i) = \mathbf{f}(i) - \overrightarrow{\mu}; i = 1, .., n\}$.

(2) Initialize the set of vertices $E = \{\mathbf{e}_1\}$ with a randomly picked sample. Initialize the set of lattice independent binary signatures $X = \{\mathbf{x}_1\} = \{(e^1_k > 0; k = 1, .., d)\}$

(3) Construct the AMM's based on the lattice independent binary signatures: M_{XX} and W_{XX}.

(4) For each pixel $\mathbf{f}^c(i)$

  (a) compute the noise corrections sign vectors$\mathbf{f}^+(i) = (\mathbf{f}^c(i) + \alpha\overrightarrow{\sigma} > \mathbf{0})$ and $\mathbf{f}^-(i) = (\mathbf{f}^c(i) - \alpha\overrightarrow{\sigma} > \mathbf{0})$

  (b) compute $y^+ = M_{XX} \boxslash \mathbf{f}^+(i)$

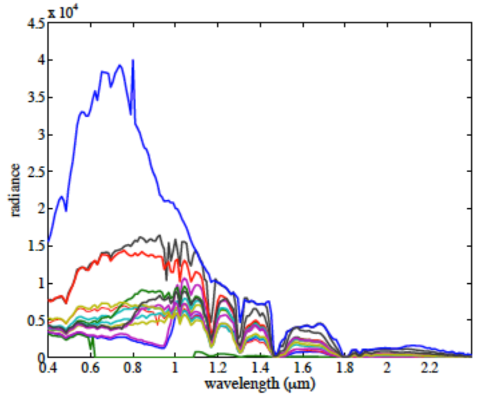
  (c) compute $y^- = W_{XX} \boxtimes \mathbf{f}^-(i)$

  (d) if $y^+ \notin X$ or $y^- \notin X$ then $\mathbf{f}^c(i)$ is a new vertex to be added to E, execute once 3 with the new E and resume the exploration of the data sample.

  (e) if $y^+ \in X$ and $\mathbf{f}^c(i) > \mathbf{e}_{y^+}$ the pixel spectral signature is more extreme than the stored vertex, then substitute $\mathbf{e}_{y^+}$ with $\mathbf{f}^c(i)$.

  (f) if $y^- \in X$ and $\mathbf{f}^c(i) < \mathbf{e}_{y^-}$ the new data point is more extreme than the stored vertex, then substitute $\mathbf{e}_{y^-}$ with $\mathbf{f}^c(i)$.

(5) The final set of endmembers is the set of original data vectors $\mathbf{f}(i)$ corresponding to the sign vectors selected as members of E.



(a) (b)

Figure : (a) patch of washington dc image, (c) EIHA endmembers

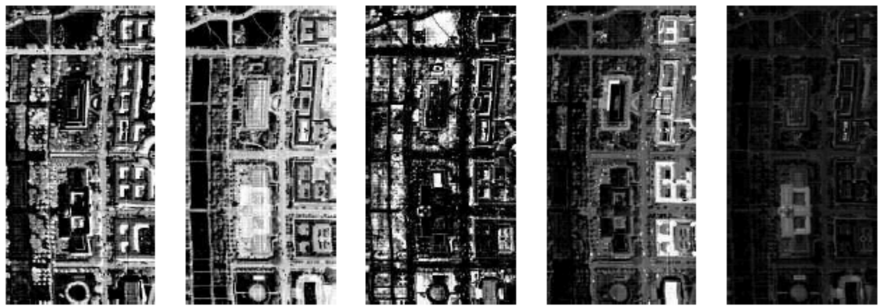


Figure : LSU estimated abundances from Washington DC patch

Multivariate ordering

Definition

A h-ordering is defined by a surjective map of the original partially ordered set onto a complete lattice $h : X \to \mathbb{L}$,

- The order in $\mathbb{L}$ induces a total order in X,

$$r \leq_h r' \Leftrightarrow h(r) \leq h(r') \quad (3)$$

Definition

Supervised h-ordering the mapping is built by supervised classification

- satisfying $h(b) = \bot$, $\forall b \in B$, and $h(f) = \top$, $\forall f \in F$,
- for background and foreground $B, F \subset X$, $B \cap F = \emptyset$,
  - $\bot$ and $\top$ are the bottom and top elements of $\mathbb{L}$

Supervised erosion and dilation

Definition

The supervised h-erosion by structural object S is

$$\varepsilon_{h,S}(I)(p) = I(q) \text{ s.t. } I(q) = \bigwedge_h \{I(s); s \in S_p\}$$

Definition

The supervised h-dilation by structural object S is

$$\delta_{h,S}(I)(p) = I(q) \text{ s.t. } I(q) = \bigvee_h \{I(s); s \in S_p\}$$

where $\bigwedge_h$ and $\bigvee_h$ are the infimum and supremum defined by the reduced ordering $\leq_h$

LAAM h-function

Definition

Given $\mathbf{c} \in \mathbb{R}^n$ and $X = \{\mathbf{x}_i\}_{i=1}^K$, $\mathbf{x}_i \in \mathbb{R}^n$; the LAAM based h_X-function is

$$h_X(\mathbf{c}) = \zeta\left(\mathbf{x}^\#, \mathbf{c}\right), \tag{4}$$

- $\mathbf{x}^\# \in \mathbb{R}^n$ is a LAAM recall result

$$\mathbf{x}^\# = M_{xx} \boxed{\wedge} \mathbf{c}$$

or

$$\mathbf{x}^\# = W_{xx} \boxed{\vee} \mathbf{c}$$

- $\zeta(\mathbf{a}, \mathbf{b})$ is the Chebyshev distance $\zeta(\mathbf{a}, \mathbf{b}) = \bigvee_i |a_i - b_i|$.

One-side ordering

Definition

one-side LAAM-supervised ordering:

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n,\ \mathbf{x} \leq_X \mathbf{y} \Longleftrightarrow h_X(\mathbf{x}) \leq h_X(\mathbf{y}). \quad (5)$$

- $h_X : \mathbb{R}^n \to \mathbb{L}_X$, where $\mathbb{L}_X = (\mathbb{R}_0^+, <)$, $\perp_X = 0$
- the Background set B s.t. $h_X(\mathbf{b}) = \perp_X = 0$
  - is the set of fixed points of the LAAM $B = \mathcal{F}(X)$

B/F ordering

Definition

The relative background/foreground supervised LAAM h-function:

$$h_r(\mathbf{c}) = h_F(\mathbf{c}) - h_B(\mathbf{c}), \quad (6)$$

Given training sets B and F

Definition

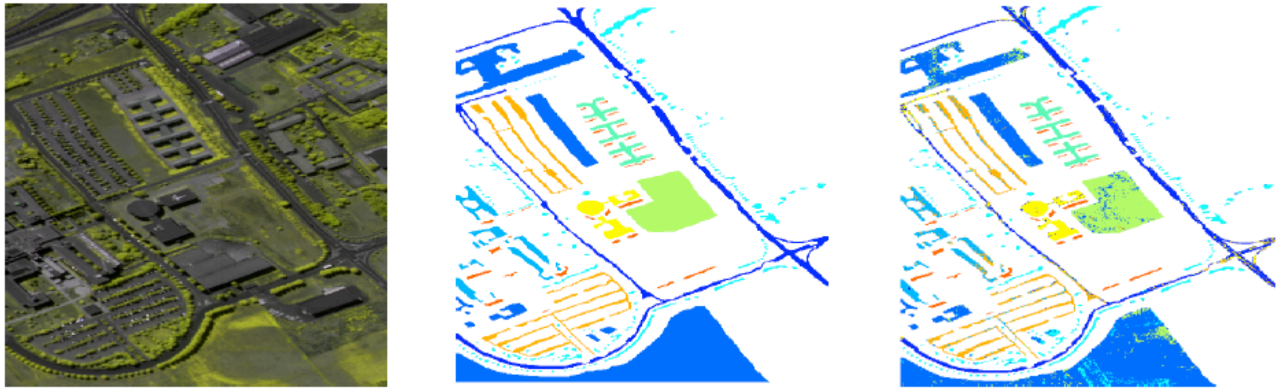
relative LAAM-supervised ordering denoted $\leq_r$:

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n,\ \mathbf{x} \leq_r \mathbf{y} \Longleftrightarrow h_r(\mathbf{x}) \leq h_r(\mathbf{y}) \quad (7)$$

Hyperspectral image spectral-spatial classification

- Independent SVM spectral classification per pixel
- Multivariate mathematical morphology provide the spatial information
  - Watershed regions from morphological gradient
    - assume homogeneous class inside each region
  - Spatial correction of SVM results

Hyperspectral image and baseline SVM classification



(a) (b) (c)

Figure : (a) Pavia image, (b) ground truth, (c) pixelwise SVM classification

Supervised morphological gradient

Definition

The h-supervised morphological gradient:

$$g_{h,S}(I) = h\left(\delta_{h,S}(I)\right) - h\left(\varepsilon_{h,S}(I)\right),$$

where $\varepsilon_{h,S}(I)$ and $\delta_{h,S}(I)$ are the h-supervised erosion and dilation

Unsupervised selection of LAAM training data

- An EIA induces a set of endmembers $E = \{\mathbf{e}_i\}_{i=1}^{p}$. Compute $D = [d_{i,j}]_{i,j=1}^{p}$, where $d_{ij} = |\mathbf{e}_i, \mathbf{e}_j|$
- One-side h-supervised ordering
  - $X = \{\mathbf{e}_{k^*} \in E\}$ such that $k^* = \arg\min_k \left\{ \frac{1}{p-1} \sum_{i \neq k} d_{ik} \right\}_{i=1}^{p}$.
- Background/Foreground h-supervised orderings
  - $F = \{\mathbf{e}_{i^*} \in E\}$ and $B = \{\mathbf{e}_{j^*} \in E\}$ such that $(i^*, j^*) = \arg\max_{i,j} \{(d_{ij})\}$

Morphological gradient results

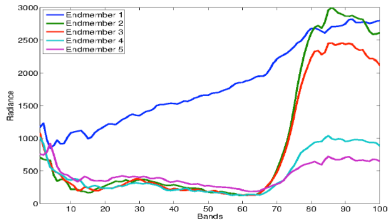


Figure : Endmembers found in the hyperspectral image

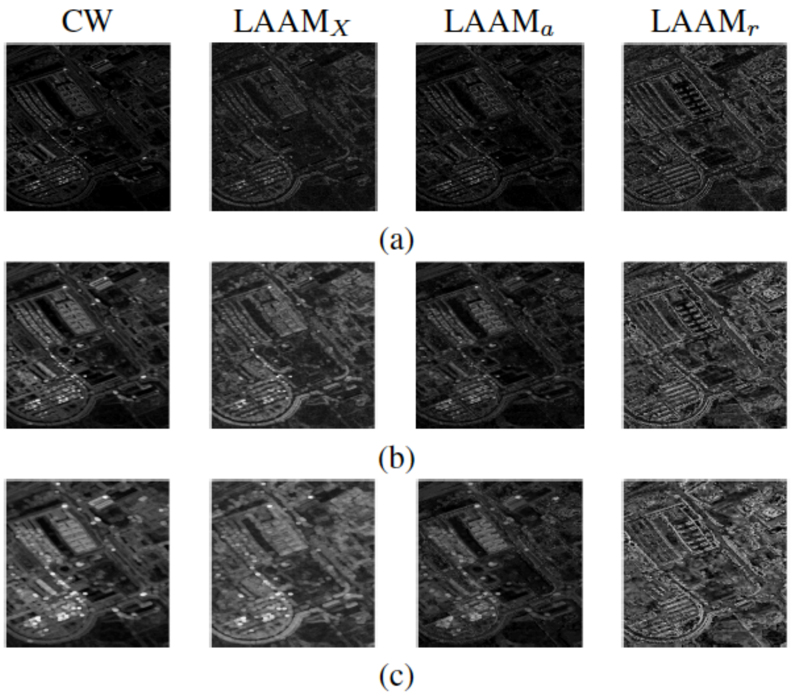


Figure : Morphological gradients with increasing structural element size

Classification results

Method		OA	AA	κ
Pixel-wise SVM		88.97	91.60	0.8565
SVM + NWHED	CW	93.41	94.39	0.9135
	LAAM_{X}	93.65	94.72	0.9167
	LAAM_{r}	92.61	93.84	0.9034
SVM+WHED	CW	95.46	95.86	0.9403
	LAAM_{X}	95.27	96.11	0.9378
	LAAM_{r}	94.91	95.71	0.9332

Table : Classification results of the Pavia University hyperspectral image: OA, AA, and Kappa (κ) values. Morphological structural element disc shaped of radius $r = 5$.

Conclusions

Conclusions

- Lattice computing defined as computing on the lattice algebra $(R, \wedge, \vee, +)$ has been maintaining its appeal in the last fifteen years.

Conclusions

- Application of lattice theory leads to new computational paradigms arising from
  - Fusion of established paradigms
    - Mathematical morphology and fuzzy systems
    - Neural networks and fuzzy systems
  - Generalization of approaches
    - Fuzzy Lattice Neurocomputing
  - Direct innovative applications
    - Feature extraction based on linear unmixing based on the identification of endmembers in the data set.

Future

- Lattice Computing may benefit from
  - Advances in random search
  - Sparsity approaches.
- A wide open field for mathematical research
- Need of open source libraries for dissemination

Thank you for your attention