Multivariate and Supervised Approaches for Mathematical Morphology in Remote Sensing

Sébastien Lefèvre

Image Sciences, Computer Sciences and Remote Sensing Laboratory (LSIIT)
Models, Image and Vision Team (MIV)
Discrete Geometry and Mathematical Morphology Group (GDMM)

University Louis Pasteur - CNRS, Strasbourg / France
lefevre@lsiit.u-strasbg.fr

JRC, Ispra
February 28th, 2008
Mathematical Morphology is a powerful toolbox for image analysis but . . .

How to extend it to multivariate data such as colour/multispectral/hyperspectral images?

How to involve some domain knowledge into the morphological processes?
Outline

1. Introduction
2. Multivariate MM
3. Supervised MM
4. Applications in Remote Sensing
5. Conclusion
Basics of Mathematical morphology

**Definition**

A framework for **analysis of spatial structures in images**. A set of operators applied on an image \( f \) with a pattern (SE) \( B \). Two fundamentals operators (erosion \( \varepsilon \) and dilation \( \delta \)) ... from which are built more complex ones.

**Theoretical framework**

For binary images, MM may be formalized with set theory ... but not for more complex images (grayscale, multispectral). Thus MM is rather defined using **complete lattice theory**.
Basics of Mathematical morphology

Complete lattice theory

A complete lattice is defined from:
- a partially **ordered set** \((L, \geq)\) (e.g. the natural order of scalars)
- an **infimum** or greatest lower bound \(\land\) (e.g. minimum)
- a **supremum** or least upper bound \(\lor\) (e.g. maximum)

Main operators

<table>
<thead>
<tr>
<th>Operator</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erosion</td>
<td>(\varepsilon_B(f)(x, y) = \land_{(r,s)\in B} f(x + r, y + s), (x, y) \in E)</td>
</tr>
<tr>
<td>Dilation</td>
<td>(\delta_B(f)(x, y) = \lor_{(r,s)\in B} f(x - r, y - s), (x, y) \in E)</td>
</tr>
<tr>
<td>Opening</td>
<td>(\gamma_B(f) = \delta_B(\varepsilon_B(f)))</td>
</tr>
<tr>
<td>Closing</td>
<td>(\phi_B(f) = \varepsilon_B(\delta_B(f)))</td>
</tr>
</tbody>
</table>
Orderings: Definitions and Properties

A binary relation $\mathcal{R}$ on a set $S$ is

- **reflexive** if $x \mathcal{R} x$, $\forall x \in S$
- **anti-symmetric** if $x \mathcal{R} y$ and $y \mathcal{R} x \Rightarrow x = y$, $\forall x, y \in S$
- **transitive** if $x \mathcal{R} y$ and $y \mathcal{R} w \Rightarrow x \mathcal{R} w$, $\forall x, y, w \in S$
- **total** if $x \mathcal{R} y$ or $y \mathcal{R} x$, $\forall x, y \in S$

An **ordering** is a reflexive, anti-symmetric and transitive relation. It is:

- a **pre-ordering** if the anti-symmetry statement does not hold
- a **partial** ordering if the totality statement does not hold
- a **total** ordering if the totality statement holds
Outline

1. Introduction

2. Multivariate MM
   - Principles
   - Partial orderings
   - Total pre-orderings
   - Total orderings

3. Supervised MM

4. Applications in Remote Sensing

5. Conclusion
**Principles**

**MM for multivariate data**
- multivalued images: \( f: \mathbb{Z}^2 \rightarrow \mathbb{R}^n, \ n > 1 \)
- spectral signature: \( f(x, y) = [f_1(x, y), \ldots, f_n(x, y)] \)

**Main operators**

- Opening:  \( \gamma_B^v(f) = \delta_B^v(\varepsilon_B^v(f)) \)
- Closing:  \( \phi_B^v(f) = \varepsilon_B^v(\delta_B^v(f)) \)

- Erosion:  \( \varepsilon_B^v(f)(x, y) = \inf_{(r,s) \in B} \{ f(x + r, y + s) \} \)

- Dilation:  \( \delta_B^v(f)(x, y) = \sup_{(r,s) \in B} \{ f(x - r, y - s) \} \)

... but there is no universal ordering relation between vectors!

E.g. min or max of \(((3, 1), (5, 5), (4, 2), (2, 4), (1, 1)) = ? \)
Barnett taxonomy of vectorial orderings (1976)

4 classes of orderings

**Marginal:** data are ordered independently inside each band
\[ \forall \mathbf{v}, \mathbf{v}' \in \mathbb{R}^n, \mathbf{v} \leq \mathbf{v}' \iff \forall i \in \{1, \ldots, n\}, \ v_i \leq v_i' \]

**Conditional:** data are ordered according to an ordered (sub)set of available bands (e.g. lexicographical ordering).

**Partial:** data are split into groups, and then ordered inside each group but not between different groups.

**Reduced:** data are reduced to scalar values and then ordered according to these values:
\[ \forall \mathbf{v}, \mathbf{v}' \in \mathbb{R}^n, \mathbf{v} \leq \mathbf{v}' \iff h(\mathbf{v}) \leq h(\mathbf{v}'), \ h : \mathbb{R}^n \rightarrow \mathbb{R} \]
Processing strategies

**Marginal processing**

Pixel = scalar value, processed independently in each band.

+ : no need to adapt existing algorithms.
- : correlation between band is ignored.
- : input vectors are not preserved.

**Vectorial processing**

Pixel = vector, processed simultaneously in all bands.

+ : correlation between band is taken into account.
- : need to adapt existing algorithms.
- : may be computationally prohibitive
Partial orderings: marginal approach

Marginal approach

It is a variant of vectorial approach using a partial ordering.

Erosion

$$\varepsilon_b^v(f)(x) = [\varepsilon_b(f_1)(x), \ldots, \varepsilon_b(f_n)(x)]^T$$

Dilation

$$\delta_b^v(f)(x) = [\delta_b(f_1)(x), \ldots, \delta_b(f_n)(x)]^T$$
Partial orderings: matrix approach

Generalisation of the marginal approach

Each image $\mathbf{F}$ is processed with a matrix of structuring elements

$\mathbf{B} = \{B_{ij}, \ i, j \in \{1, \ldots, n\}\}$.

Equivalent to marginal approach if:

$\mathbf{B} = \{B_{ij} \mid (B_{ij} = \emptyset, \ i \neq j) \land B_{ii} = b, \ i, j \in \{1, \ldots, n\}\}$

Erosion

$\varepsilon_{m, \mathbf{B}}(\mathbf{F}) = (E_1, \ldots, E_n)$ with $E_i = \bigwedge_{j=1}^{n} \varepsilon_{B_{ji}}(F_j), \ i \in \{1, \ldots, n\}$

Dilation

$\delta_{m, \mathbf{B}}(\mathbf{F}) = (D_1, \ldots, D_n)$ with $D_i = \bigvee_{j=1}^{n} \delta_{B_{ij}}(F_j), \ i \in \{1, \ldots, n\}$
Partial orderings: comments

Drawbacks of marginal approaches

Correlation between band is ignored.
Spectral composition of images may be altered.

Space transformation (decorrelation)

- Principal Component Analysis
- Discrete Cosine Transform
- Perceptual colour space

Corrected marginal approach

Computed values can be replaced by the closest existing values in the input image, thus input vectors are preserved.
Total pre-orderings

Reduced approaches

- ordering based on a cumulative distance
  \[ \forall \mathbf{v}_k, \mathbf{v}_l \in \{\mathbf{v}_j\}, \quad \mathbf{v}_k \leq \mathbf{v}_l \iff \sum_j d(\mathbf{v}_k, \mathbf{v}_j) \leq \sum_j d(\mathbf{v}_l, \mathbf{v}_j) \]
- ordering based on a distance to a reference vector
  \[ \forall \mathbf{v}_k, \mathbf{v}_l \in \{\mathbf{v}_j\}, \quad \mathbf{v}_k \leq \mathbf{v}_l \iff d(\mathbf{v}_k, \mathbf{v}_{\text{ref}}) \leq d(\mathbf{v}_l, \mathbf{v}_{\text{ref}}) \]
- ordering based on a double distance
  \[ h(\mathbf{v}) = w(\mathbf{v}, \mathbf{v}_c) \cdot a(\mathbf{v}, \mathbf{v}_c) + (1 - w(\mathbf{v}, \mathbf{v}_c)) \cdot d_E(\mathbf{v}, \mathbf{v}_c) \]

Conditional approaches

- ordering on one component
  \[ \mathbf{v} \leq \mathbf{v}' \iff v_1 \leq v'_1 \]
- restricted lexicographical ordering (with \( k < n \))
  \[ \mathbf{v} <_L \mathbf{v}' \iff \exists i \in \{1, \ldots, k\}, \quad (\forall j < i, \quad v_j = v'_j) \land (v_i < v'_i) \]
Total pre-orderings: comments

**Advantages**

- Vectors are preserved
- Possible symmetry between bands (if desired).

**Drawbacks**

- Several different vectors may be considered as same extrema.
- Possible instability of the dilation operator.
Total orderings: some examples

<table>
<thead>
<tr>
<th>Lexicographical ordering</th>
<th>( \mathbf{v} &lt;_L \mathbf{v}' ) &amp; ( \exists \ i \in {1, \ldots, n}, \ (\forall \ j &lt; i, \ \mathbf{v}_j = \mathbf{v}'_j) \land (\mathbf{v}_i &lt; \mathbf{v}'_i) )</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Lexicographical ordering with additional component</th>
<th>( \mathbf{v} \leq \mathbf{v}' ) &amp; ( [|\mathbf{v}|, \mathbf{v}_1, \ldots, \mathbf{v}_n]^T \leq_L [|\mathbf{v}'|, \mathbf{v}'_1, \ldots, \mathbf{v}'_n]^T )</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Lexicographical ordering modulo ( \alpha )</th>
<th>( \mathbf{v} \leq \mathbf{v}' ) &amp; ( [\lceil \mathbf{v}_1/\alpha \rceil, \mathbf{v}_2, \ldots, \mathbf{v}_n]^T \leq_L [\lceil \mathbf{v}'_1/\alpha \rceil, \mathbf{v}'_2, \ldots, \mathbf{v}'_n]^T )</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Ordering based on bit mixing</th>
<th>Scalar representation of a vector by mixing the different bits: ( h(\mathbf{v}) = \sum_{m=1}^{k} \left{ 2^{n \cdot (k-m)} \cdot \sum_{i=1}^{n} 2^{n-i} \cdot \mathbf{v}_{i,m} \right} )</th>
</tr>
</thead>
</table>
Total orderings : comments

**Advantages**

Vectors are preserved.
Extrema are always unique.

**Drawback**

Dissymmetry since a high importance is given to first components.

**Solution**

Dissymmetry can be attenuated using a space transformation. Advanced orderings to ensure a better weight between bands.
Total orderings: some balanced approaches

\( \alpha \)-trimmed lexicographical extrema

Keep iteratively only the \( \alpha \% \) greatest/lowest values of each band. Consider normalized distance to the extrema to avoid edge effects. For a finer tuning, specific \( \alpha_i \) values can be used for each band.

Quantisation-based \( \alpha \)-lexicographical ordering

Avoid uniform quantisation and model desired priority distribution:

\[ \mathbf{v} \leq \mathbf{v}' \iff [v_1, v_2, \ldots, v_n]^T \leq_L [v'_1, v'_2, \ldots, v'_n]^T \]

where \( w_1 \) and \( w'_1 \) represents the equivalence group of \( v_1 \) and \( v'_1 \).

Marker-based lexicographical ordering

Priority distribution varies for each pixel, \( \forall x, y, s, t \in \mathbb{Z} : \)

\[ g(x, y) < g(s, t) \iff [m(x, y), g_2(x, y), \ldots, g_n(x, y)]^T \leq_L [m(s, t), g_2(s, t), \ldots, g_n(s, t)]^T \]
Outline

1. Introduction
2. Multivariate MM
3. Supervised MM
   - Domain knowledge and MM
   - Using expert knowledge
   - Using learning samples
4. Applications in Remote Sensing
5. Conclusion
Domain knowledge and MM

Existing methods for knowledge integration

- Structuring elements
  It is possible to define the shape and size of the structuring element depending on the problem under consideration. *But definition of the SE may be irrelevant for some objects!*

- Choice of morphological operators
  To solve a problem, a morphological expert will select the appropriate operators. *But the expert is not always here!*

Two kinds of knowledge

- **Expert knowledge** represented as a database, a set of rules, an ontology
- **Learning samples** or labelled visual examples
Various knowledge may be available for the object(s)

- shape(s)
- size(s)
- spectral composition(s)
- textural description(s)
- number and model of classes
- ...

They can be either:

- given and formalized by the expert
- estimated from the learning data
Using expert knowledge

An object is characterized by / a SE can be defined by

- spatial information
- spectral information
- combination of both

Thus several predefined functional SE (shape and intensity profile) can be used, one for each band of interest.

The expert knowledge may be less precise

- only the number of classes of interest is known
- only an approximate size and shape are known

In both cases, several single results are computed (for each cluster, each size, each shape, etc).
There is finally a need to merge the results (e.g. with union).
Using learning samples

Learning dataset and supervised classification process

Spectral/textural information is given by the learning dataset. This dataset is used by a fuzzy supervised classification process.

A knowledge-based representation space

A pixel is associated with membership values to each class. Vectors contain class membership values instead of spectral values. But both can also be combined (e.g. average, product).
Using learning samples

And spatial information?
Remark: the spatial location of learning samples may also be relevant and could be used.

Towards an iterative process
If posterior evaluation is possible, optimal parameters for the morphological process can be estimated:

- shape
- size
- spectral band
- other parameters (e.g. for oversegmentation reduction)
Outline

1. Introduction
2. Multivariate MM
3. Supervised MM
4. Applications in Remote Sensing
   - Pixel-based Classification using DMP
   - Object Detection using HMT
   - Image Segmentation using Watershed
5. Conclusion
Pixel-based Classification using DMP: Experimental setup

Methodology

- Computation of a DMP (using both openings and closings): \( \Delta(f)(x, y) \) (with \( \lambda = 5 \) and \( \omega = 3 \))
- Bayes supervised classification with 5 classes: buildings, water, shadow, roads, vegetation
- Learning dataset contains 1% of the full dataset, 10-fold cross-validation

Data

- VHR Image of Strasbourg (1100 × 900 pixels)
- SPOT 5 simulation, spatial resolution of 1.3m and 3 bands (G, R, NIR)
Multispectral image classification

Visual result with 5 classes:
Results

Precision

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Roads</td>
<td>83.3</td>
<td>82.4</td>
<td>85.5</td>
<td>85.5</td>
<td>85.6</td>
</tr>
<tr>
<td>Buildings</td>
<td>69.3</td>
<td>72.3</td>
<td>68.6</td>
<td>68.0</td>
<td>68.4</td>
</tr>
<tr>
<td>Water</td>
<td>91.3</td>
<td>96.8</td>
<td>89.5</td>
<td>89.7</td>
<td>89.8</td>
</tr>
<tr>
<td>Vegetation</td>
<td>94.0</td>
<td>94.4</td>
<td>93.1</td>
<td>93.0</td>
<td>93.4</td>
</tr>
<tr>
<td>Shadow</td>
<td>94.2</td>
<td>92.9</td>
<td>95.1</td>
<td>94.7</td>
<td>95.2</td>
</tr>
<tr>
<td>Overall</td>
<td>87.0</td>
<td>88.2</td>
<td>86.9</td>
<td>86.8</td>
<td>87.0</td>
</tr>
</tbody>
</table>

Comments

- Global quality is rather close for the different orderings
- Disymmetric approaches may be relevant for some classes
- Marginal approach does not preserve spectral signatures!
Object Detection using HMT: Building Detection

Assumptions

A building has a regular shape (e.g. rectangle) but it may contains several materials, thus its content may not be homogeneous but rather composed of homogeneous parts.

Convert the input image to a set of binary images through clustering and addition
Algorithm and results

Algorithm

- Image clustering and generation of binary images (clusters)
- Optimal opening of each cluster (using 2-D granulometry)
- Adaptive Hit or Miss Transform and geodesic reconstruction
- Unification of the results (and additional HMT)
Object Detection using HMT : Coastline Extraction

Extended SEs for Multispectral Hit or Miss Transform

- Size and shift from the origin
- Spectral band
- Class (foreground, background), i.e. operator to be used
- Threshold

HMT fitting

\[
\text{Fitting}_\Omega(I)(p) = \bigwedge_{s \in \Omega} \text{Fitting}_s(I)(p)
\]

\[
\text{Fitting}_s(I)(p) = \begin{cases} 
\varepsilon_{s\text{shape}}(l_{sb})(p) \geq s_t & \text{if } s_{\text{type}} = \varepsilon \\
\delta_{s\text{shape}}(l_{sb})(p) \leq s_t & \text{if } s_{\text{type}} = \delta
\end{cases}
\]
Coastline Extraction

HMT Valuation

HMT returns a value to each fitted pixel:

$$\text{Valuation}_\Omega(I)(p) = \sum_{s \in \Omega} \frac{\text{Valuation}_s(I)(p)}{|\Omega|}$$

with

$$\text{Valuation}_s(I)(p) = \begin{cases} 
\varepsilon_{\text{shape}}(I_{sb})(p) - s_t \\
I_{sb}^+ - s_t \\
I_{sb}^- - s_t - \gamma_{\text{shape}}(I_{sb})(p) \\
\gamma_{\text{shape}}(I_{sb})(p) - s_t - I_{sb}^-
\end{cases}$$

if $s_{\text{type}} = \varepsilon$

if $s_{\text{type}} = \delta$

where $[I_k^-, I_k^+]$ is the range of pixel values for the band $I_k$. 

Sébastien Lefèvre
LSIIT, ULP-CNRS
Result on a VHR Image

\[
\begin{align*}
\{ s_{\text{shape}} &= 180 + 0, \quad s_b = \text{NDVI}, \quad s_t = 0, \quad s_{\text{type}} = \delta \}, \\
\{ s_{\text{shape}} &= 800 + 300, \quad s_b = \text{NIR}, \quad s_t = 0, 125, \quad s_{\text{type}} = \delta \}, \\
\{ s_{\text{shape}} &= -180 + 0, \quad s_b = \text{NDVI}, \quad s_t = 0, 004, \quad s_{\text{type}} = \varepsilon \}
\end{align*}
\]
Comparison with other approaches
Watershed Segmentation of urban areas

Context

A set of labelled samples is given by the expert.
Segmentation of urban areas

Algorithm

- Conversion from the input space to the membership space
- Compute watershed relief from gradient of membership values
- Perform watershed algorithm
- Evaluate results through classification to optimize oversegmentation reduction parameters.

Memberships maps
1 map per class of interest
Segmentation of urban areas

Results

unsupervised watershed

optimal watershed

supervised watershed

supervised optimal watershed
Watershed Segmentation of coastal areas

Context, coastline extraction

In such a case, a segmentation into two objects may be useful.

Method

- Manual definition of the markers (used in the learning step)
- Supervised fuzzy classification to build membership maps $w_i$
- Replace unsupervised relief $f$ by $f_i = (1 - w_i) \cdot f$
- Flood each marker $M_i$ using relief $f_i$
- $p$ is assigned to $M_i$ if first reached by flooding from $M_i$
Segmentation of coastal areas

input image

result with markers
Segmentation of urban areas

marker-based watershed

using spectral markers

with one marker for each building

and a single marker (bottom right) for the background
# Conclusions and Perspectives

## Issues
- Adaptation of MM to multispectral/hyperspectral images
- Knowledge integration for MM-based image analysis

## Contributions
- Survey of Multivariate MM and new vectorial orderings
- Various ways for using domain knowledge into supervised MM
- Detection, segmentation, classification in urban/coastal RS

## Perspectives
- Check the relevance of proposed orderings in the RS field
- Define new physically meaningful orderings for RS
- Combine multivariate and supervised advances
Some further readings . . .

Theoretical advances in multivariate MM

Supervised morphological segmentation methods
- Lefèvre (2007), *Knowledge from markers in watershed segmentation*, IAPR CAIP

Extensions of the HMT operator for template matching
Sébastien Lefèvre
LSIIT, University Louis Pasteur - CNRS
Strasbourg, France
lefevre@lsiit.u-strasbg.fr

Collaborators:
Erchan Aptoula, PhD Student
Sébastien Derivaux, PhD Student
Jonathan Weber, MSc Student