Spatial & Spectral Morphological Template Matching

J. Weber[a,*], S. Lefèvre[b]

[a]*Université de Lorraine, LORIA-UMR 7503, Vandœuvre-lès-Nancy, F-54506, France*
[b]*IRISA, University of South Brittany, France*

Abstract

Template matching is a very topical issue in a wide range of imaging applications. Mathematical morphology offers the hit-or-miss transform, an operator which has been successfully applied for template matching in binary images. More recently, it has been extended to greyscale images and even to multivariate images. Nevertheless, these extensions, despite being relevant from a theoretical point-of-view, might lack practical interest due to the inherent difficulty to set up correctly the transform and its parameters (e.g. the structuring functions). In this paper, we propose a new and more intuitive operator which allows for morphological template matching in multivariate images from both a spatial and spectral point of view. We illustrate the potential of this operator in the context of remote sensing.

Keywords: Hit-or-miss transform, remote sensing, multivariate morphology, template matching.

*Corresponding author.
Phone: +33 (0) 329 536 041.
Address: LORIA Nancy, Campus Scientifique - BP 239, 54506 Vandœuvre-les-Nancy Cedex, France
Email addresses: `jonathan.weber@loria.fr` (J. Weber), `sebastien.lefevre@univ-ubs.fr` (S. Lefèvre)

1. Introduction

To face the increasing of amount, quality, size and diversity of images that are produced everyday, automatic tools for information extraction are required. Among these tools, template matching methods aim to extract relevant information from various images for a given user. Template matching is thus a very topical issue.

Among the imaging toolboxes available, mathematical morphology offers various operators with intrinsic ability to deal with spatial information. Its most famous template matching method, namely the hit-or-miss transform (HMT), has been widely used since its first definition for binary images by Matheron [11] and Serra [16]. This operator has been later extended to grey-level images, leading to several definitions which have been recently reviewed and unified by Naegel et al. [12]. We can even notice a first attempt to deal with multivariate images [2], but it relies on a vectorial ordering which is not always easy to choose among the many solutions available [1]. A more convenient way of setting up a template matching operator is needed to make this operator really useful in practice.

In a previous work [21], we have briefly introduced a new definition of the hit-or-miss transform for multivariate images and illustrated its potential as a template matching operator by a first study in remote sensing (related to coastline extraction). We extend here this preliminary work, by giving much more theoretical background, comparing the proposed approach with existing works, addressing the problem of efficiency. Moreover, we also propose some applications which illustrate the relevance of the proposed technique for morphological template matching in multivariate images and we compare it to common template matching methods.

This paper is organised as follows. In the next section, we recall existing definitions of the hit-or-miss transform, which allow morphological template matching on binary, greyscale and even multivariate images. We then introduce in Sec. 3 a new definition which makes this operator easier to set up from a practical point-of-view and study this operator in the context of the theory of mathematical morphology. In Sec. 4, we explain how this operator may be used to achieve template matching in a wide range of contexts and illustrate its abilities by various experiments in remote sensing. We finally conclude in Sec. 5 and indicate some future directions.

2. A Review on Morphological Template Matching

Mathematical morphology (MM) is a theoretical framework introduced more than 40 years ago by Matheron [11] and Serra [16]. It aims to compute quantitative description of geometrical structures through spatial analysis. The rich set of operators it offers has been found very useful for solving many tasks in image processing and analysis for several decades [19].

In this section, we will review the solutions offered by mathematical morphology to solve the problem of template matching, from the earliest works related to binary images to most recent ones dealing with multivariate (e.g. multispectral) images. We will also introduce the mathematical notations which will be used throughout the rest of this paper.

2.1. Binary case

In binary images, mathematical morphology has been first formalised using the set theory. In this framework, we consider that an image contains some objects or foreground X and their complement or background X^c. Template matching consists then in searching for areas of the image where both foreground and background match predefined patterns, as illustrated in Fig. 1. Within the scope of binary mathematical morphology, these patterns are also sets and called *structuring elements* and template matching is performed through the *hit-or-miss transform*, which is defined as

$$\mathrm{HMT}_{F,G}^{\mathrm{binary}}(X) = \{\mathbf{x} \mid (F)_{\mathbf{x}} \subseteq X, (G)_{\mathbf{x}} \subseteq X^c\} \tag{1}$$

where $(\cdot)_{\mathbf{x}}$ denotes the translation by $\mathbf{x}$, while F and G represent respectively the foreground and background structuring elements, with the condition $F \cap G = \emptyset$.

In other words, the hit-or-miss transform returns pixels $\mathbf{x}$ such that both F and G, when centered in $\mathbf{x}$ (i.e., $(F)_\mathbf{x}$ and $(G)_\mathbf{x}$) match foreground X and background X^c, respectively.

This operator can also be written as an intersection of two erosions, one for the foreground (i.e. the *hit*) and one for the background (i.e. the *miss*). Since erosion ε and dilation δ, the two elementary morphological operators, are dual with respect to complementation (i.e. $\varepsilon_F(X) = (\delta_{\breve{F}}(X^c))^c$), the hit-or-miss tranform can even be written using only foreground pixels:

$$\begin{aligned}\mathrm{HMT}^{\mathrm{binary}}_{F,G}(X) =& \varepsilon_F(X) \cap \varepsilon_G(X^c) \\ =& \varepsilon_F(X) \cap \left(\delta_{\breve{G}}(X)\right)^c \\ =& \varepsilon_F(X) \setminus \delta_{\breve{G}}(X)\end{aligned} \tag{2}$$

with $\breve{G}$ denoting the reflection, i.e. $\breve{G} = \{-x \mid x \in G\}$. We recall that erosion and dilation are defined in the binary case as:

$$\varepsilon_F(X) = \bigcap_{\mathbf{b}\in B} (X)_{-\mathbf{b}} \tag{3}$$

$$\delta_F(X) = \bigcup_{\mathbf{b}\in B} (X)_{\mathbf{b}} \tag{4}$$

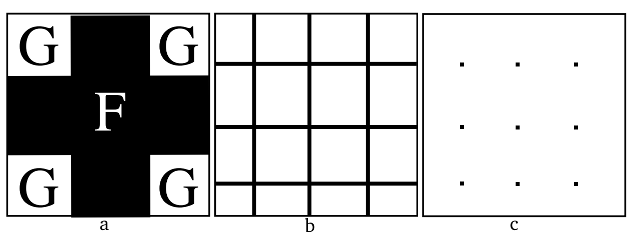


Figure 1: Binary HMT: a) zoom on the couple of SEs (foreground and background); b) binary lattice; c) intersections matched by the HMT.

2.2. Greyscale case

Within greyscale images, images are not composed of foreground and background anymore. Thus, set theory was replaced with lattice theory to design morphological operators [17]. We consider from now that an image is represented as a function $w : E \to T$, where $E \subset \mathbb{N}^2$ denotes the spatial domain on which pixel coordinates $\mathbf{p} = (x, y)$ are defined and T being the value domain representing pixel values v. Thus the image w aims to assign a value $v \in T$ to every pixel $\mathbf{p} \in E$.

Similarly to images, structuring elements may now be defined as functions (sometimes called *structuring functions*) $f : D \to T$, with D their spatial domain. But let us observe that very often, even with greyscale images, only flat structuring elements are used. These structuring elements are defined as sets, leading to an uniform processing of the neighbourhood of each pixel. Such flat structuring elements are particular cases of structuring functions, where pixel values are set to the additive identity of the value space (most often zero), e.g. for a structuring element F we have the following structuring function: $\forall\ \mathbf{p} \in F, f(\mathbf{p}) = 0$.

To perform morphological template matching in greyscale images, the hit-or-miss transform has to be defined on such images. This may be achieved by extending the previous definitions given for binary images, in particular Eq. (2) where image complementation (which may be hard to define on greyscale images) is not necessary:

$$\mathrm{HMT}^{\mathrm{grey}}_{F,G}(w) = \varepsilon_F(w) - \delta_{\check{G}}(w) \tag{5}$$

but some pixels may then be assigned negative values, i.e. when $\varepsilon_F(w)$ is lower than $\delta_{\check{G}}(w)$. To avoid such a case, all pixels where $\varepsilon_F(w) < \delta_{\check{G}}(w)$ may be set to 0, as proposed by Soille [18]:

$$\begin{aligned}\mathrm{HMT}^{\mathrm{Soille}}_{F,G}(w)(\mathbf{p}) &= \begin{cases} \varepsilon_F(w)(\mathbf{p}) - \delta_{\check{G}}(w)(\mathbf{p}) & \text{if } \varepsilon_F(w)(\mathbf{p}) \geq \delta_{\check{G}}(w)(\mathbf{p}) \\ 0 & \text{otherwise} \end{cases} \\ &= \max\{\varepsilon_F(w)(\mathbf{p}) - \delta_{\check{G}}(w)(\mathbf{p}), 0\}\end{aligned} \tag{6}$$

We recall that erosion and dilation are defined in the greyscale case as:

$$\varepsilon_f(w)(\mathbf{p}) = \inf_{\mathbf{q}\in \mathrm{supp}(f)} (w(\mathbf{p}+\mathbf{q}) - f(\mathbf{q})) \tag{7}$$

$$\delta_f(w)(\mathbf{p}) = \sup_{\mathbf{q}\in \mathrm{supp}(f)} (w(\mathbf{p}-\mathbf{q}) + f(\mathbf{q})) \tag{8}$$

with $\mathrm{supp}(f)$ denoting the support of function f.

Since greyscale images enable the use of structuring functions, more complex definitions may be given, as by Ronse [15]:

$$\mathrm{HMT}^{\mathrm{Ronse}}_{f,g}(w)(\mathbf{p}) = \begin{cases} \varepsilon_f(w)(\mathbf{p}) & \text{if } \varepsilon_f(w)(\mathbf{p}) \geq \delta_{g^*}(w)(\mathbf{p}) \\ \perp & \text{otherwise} \end{cases} \tag{9}$$

where g^* denotes the dual of g, i.e. $g^*(\mathbf{p}) = -g(-\mathbf{p})$ and $\perp$ denotes a specific value (e.g., $-\infty$ or 0) as it will be explained next. Apart from the use of structuring functions (which leads to duality rather than reflection of the background pattern), this definition also suggests not to use the difference between erosion and dilation but rather the erosion itself as the result in each pixel matched by the pair of structuring elements, as illustrated by Fig. 2. Nevertheless, these two definitions share some common properties and behaviors, as pointed out by Naegel et al. [12] who reviewed existing HMT definitions for greyscale images and proposed a unified formulation.

In this unified formulation, the hit-or-miss transform is performed in two steps. First, a *fitting* occurs to check if a given pixel matches the pattern defined by the pair of structuring elements. This process shall ensure the translation invariance property, i.e. the fitting is the same for w and all its possible (vertical or greylevel) translations w_t (with $w_t(\mathbf{p}) = w(\mathbf{p}) + t$). The second step, where each matched pixel is given a matching score as value (while unmatched pixels are set to $\perp$), is called the *valuation* step. Contrary to binary images where all matched pixels are set to foreground while unmatched pixels are set to background, here the problem of value assignment of matched pixels is not trivial and does not lead to a unique proposition. Indeed, while the definition from Soille aims to measure how well a given image area fit the pair of structuring elements, the definition from Ronse makes possible to design morphological filters based on HMT (e.g. HMT opening is obtained by dilating the HMT result).

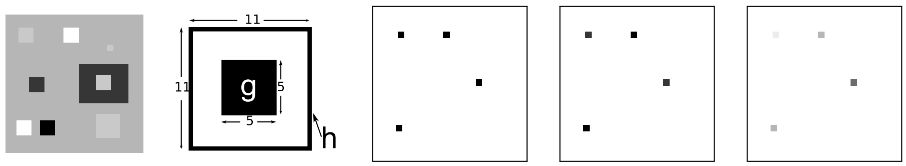


Figure 2: Greyscale HMT : a) original image (64 × 64 pixels); b) couple of structuring elements/functions used within the HMT; c) fitting result; valuation result with Ronse's (d) and Soille's (e) definition (results are displayed in inverse grey levels and for the sake of clarity, $\perp$ value was replaced with 0).

Let us also note that other attempts have been considered to define greyscale template matching within mathematical morphology [4]. However, we will not review in further details these approaches since we are concerned with multivariate template matching.

2.3. Multivariate case

Compared to greyscale images, multivariate images bring a new dimension related to spectral information. Indeed, in such images, the value space is not monodimensional anymore but rather defined as a set of spectral dimensions $\mathbf{T} = (T_b)_{b \in [1,n]}$ with T_b being the value domain related to the color or spectral band b. A multivariate image can then be written as a function $\mathbf{w} : E \to \mathbf{T}$, where each pixel $\mathbf{p}$ is assigned a set of spectral values $\mathbf{v} = (v_1, \ldots v_n)$ for a n-band image. In other words, a multivariate image is built from a set of greyscale images, i.e. $\mathbf{w} = (w_1, \ldots, w_n)$. If spectral bands are comparable (i.e. they have the same definition domain), we can also represent a multivariate image by a function $w : E, B \to T$ with B the set of spectral bands. Then, for each spectral band b, a pixel $\mathbf{p}$ is assigned a greyscale value, i.e. $w(\mathbf{p}, b) = v_b$.

These two image representations are complementary and may be used to formulate the hit-or-miss transform (as well as other morphological operators) for multivariate images. The easiest way to do so is to consider a multivariate image $\mathbf{w}$ as a set of greyscale images w_b and to apply the hit-or-miss transform independently on each of these images. This approach is known as a marginal strategy and may be written:

$$\mathrm{HMT}^{\mathrm{marginal}}_{F,G}(\mathbf{w}) = \Big(\mathrm{HMT}^{\mathrm{grey}}_{F,G}(w_1),\ \ldots,\ \mathrm{HMT}^{\mathrm{grey}}_{F,G}(w_n)\Big) \tag{10}$$

with $\mathrm{HMT}^{\mathrm{grey}}$ denoting any greyscale definition of the hit-or-miss transform recalled in the previous section. Despite its relative simplicity, the marginal strategy presents several drawbacks. First, due to the independent processing of each spectral band, it completely ignores the correlated information which may be spread on the different bands. Thus the template matching process analyses every pixel without taking into account the full spectral information (or color) it contains. Moreover, the processing of each band is usually done considering the same set-up (i.e. a unique pair of structuring elements or functions are used for all spectral bands), while the expected operator behavior may depend on the spectral content of the image. Finally, there is no guarantee of vector preservation since values which were not in the input image may appear in the resulting image, thus producing new spectral signatures. Let us note however that, even if this may be a major issue for many image processing tasks, template matching does usually not suffer from this disadvantage except if the hit-or-miss transform is used in a morphological filtering context.

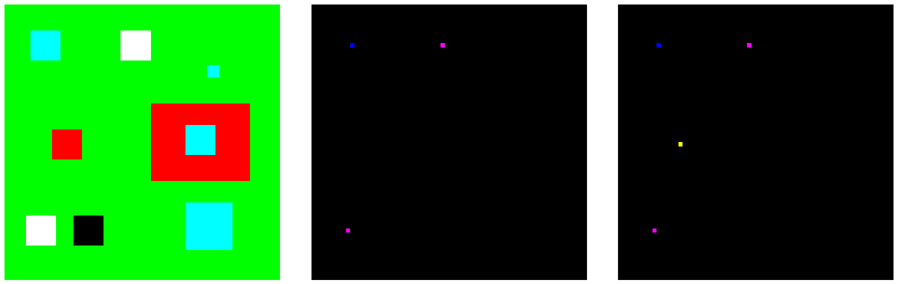
To avoid the drawbacks induced by the marginal strategy, the vectorial strategy may be involved. It requires selecting a given vectorial ordering which will be used to determine the extrema within a set of vectors representing pixel values. While greyscale images contain scalar values which are easily comparable using a unique natural ordering, there are plenty of vectorial orderings available to (morphologically) process multivariate images and a recent survey has been provided by Aptoula and Lefèvre [1]. Thus vectorial hit-or-miss transform has been subsequently addressed by Aptoula et al. [2] and it has been shown that, to apply morphological template matching on multivariate images using the hit-or-miss transform, the only requirement is to involve a vectorial ordering $\leq_v$ which is translation invariant (i.e. $\forall\, \mathbf{v}, \mathbf{v}', \mathbf{t} \in \mathbf{T},\ \ \mathbf{v} \leq_v \mathbf{v}' \ \ \Leftrightarrow\ \ \mathbf{v} + \mathbf{t} \leq_v \mathbf{v}' + \mathbf{t}$). Under this assumption, the vectorial hit-or-miss transform is a simple extension of greyscale formulations, e.g. for Ronse and Soille definitions:

$$\mathrm{HMT}^{\mathrm{Soille}}_{\mathbf{F},\mathbf{G}}(\mathbf{w})(\mathbf{p}) = \begin{cases} \left\| \varepsilon_{\mathbf{F}}(\mathbf{w})(\mathbf{p}) - \delta_{\breve{\mathbf{G}}}(\mathbf{w})(\mathbf{p}) \right\| & \text{if } \varepsilon_{\mathbf{F}}(\mathbf{w})(\mathbf{p}) \geq_v \delta_{\breve{\mathbf{G}}}(\mathbf{w})(\mathbf{p}) \\ \perp & \text{otherwise} \end{cases} \tag{11}$$

$$\mathrm{HMT}^{\mathrm{Ronse}}_{\mathbf{f},\mathbf{g}}(\mathbf{w})(\mathbf{p}) = \begin{cases} \varepsilon_{\mathbf{f}}(\mathbf{w})(\mathbf{p}) & \text{if } \varepsilon_{\mathbf{f}}(\mathbf{w})(\mathbf{p}) \geq_v \delta_{\mathbf{g}^*}(\mathbf{w})(\mathbf{p}) \\ \perp & \text{otherwise} \end{cases} \tag{12}$$

where $(\mathbf{F}, \mathbf{G})$ and $(\mathbf{f}, \mathbf{g})$ respectively denote pairs of multivariate flat structuring elements and multivariate structuring functions, which can nevertheless be replaced with their univariate alternative if we assume a uniform behavior of the operator on all spectral bands. Let us note that, in the multivariate expression of the Soille operator, the Euclidean norm $\|\cdot\|$ is used only as an illustrative vector distance measure and any other measure may be involved. Depending on the vectorial ordering involved in the comparison process, one can ensure both usage of correlated information and preservation of input vectors. But this strategy also needs to select an adequate vectorial ordering, which may be tedious since a wide range of solutions are available [1]. Moreover, the behavior of a given vectorial ordering, with respect to a multivariate image and a pair of multivariate structuring elements/functions, is still to be understood. To illustrate, Fig. 3 shows results obtained with two possible extensions of HMT to multivariate data, considering a 5×5 square foreground SE surrounded by a 1 pixel thin frame background SE. While most of the square template are extracted with both methods, we can notice that marginal approach (left) fails to identify the red square surrounded by a green background, conversely to the vectorial approach based on lexicographical ordering (right). Let us however precise that in the multivariate case, results strongly depend on the vectorial ordering and might be hard to predict.

More recently, Velasco-Forero and Angulo [20] have introduced an original



(a) (b) (c)

Figure 3: a) original image (256×256 pixels) ; b) result of a marginal HMT (with a 5×5 square foreground SE and a 1 pixel thin frame background SE) ; c) result of a vectorial HMT using lexicographical ordering and the same SE.

solution where user knowledge (i.e. learning samples) is used to learn the vectorial ordering to be used within the HMT operator. By this way, there is no need to select carefully a given vectorial ordering. Indeed, a reduced ordering (sorting pixel values from background to foreground) is automatically inferred from learning samples and is subsequently used in the HMT operator. Let us note however that in some cases, a clear distinction between foreground and background spectral values is not so easy to obtain, thus preventing us from relying on such an ordering strategy.

We believe that multivariate definition of the hit-or-miss transform may still be explored to propose intuitive tools to be used in practice. Thus, we introduce in next section a new operator which allows to perform both spatial and spectral template matching.

3. Spatial & Spectral Morphological Template Matching

In the previous section, we have recalled existing definitions of the hit-or-miss transform for morphological template matching, from the early case of binary images to more recent propositions related to greyscale and even multivariate images. As we will discuss in this section, despite their theoretical interest, these works still have to show their practical interest for morphological template matching in real scenarios when both spatial and spectral information have to be taken into account. So we propose here a new and rather intuitive definition of the hit-or-miss transform, which aims to be easily involved in a morphological template matching process (as it will be illustrated in the next section). We will also show that this original definition has its foundations within the standard morphological framework. After having pointed out the differences between our proposition and related works, we will finally discuss implementation issues.

3.1. On the practicability of existing HMT definitions for template matching

In the previous section, we have recalled the main proposed approaches for template matching within the scope of mathematical morphology. While first attempts were related to binary images, several authors have recently proposed some extensions of the hit-or-miss transform to deal with greyscale or even multivariate images. However, these definitions, despite being theoretically sound, lack practical interest when real-life applications of template matching in multivariate (e.g. color or multispectral) images are being considered.

Indeed, a first problem comes when template matching needs to rely on color or spectral properties. While the marginal strategy is unable to take care of the correlated information, adopting a vectorial strategy would require to select the adequate vectorial ordering scheme for the problem under consideration, which may be very tedious.

Moreover, since the hit-or-miss definition relies on only two structuring elements (whatever the kind of images used: binary, greyscale, multivariate), it may be unsuitable when dealing with complex templates. Thus, if the object or feature to be matched is non-uniform and composed of several (i.e. more than 2) parts with different intensity or spectral properties (e.g. plane with white fuselage, black wings parked on a grey airway), the hit-or-miss transform will very probably fail to achieve the template matching task.

Another source of failure for the hit-or-miss transform when used in the context of template matching is related to its intrinsic behavior. Since the fitting step relies on the difference between the erosion and dilation results, it can only achieve contrast-based template matching. When absolute intensity or spectral information have to be taken into account in the template matching process, existing definitions are not relevant anymore. Indeed, in some situations, prior knowledge on the minimal or maximal spectral or intensity values of the sought object may be available and they are worth being exploited. It is especially true in the field of remote sensing where the spectral behaviour of the objects under study may be well known.

All these limitations, when coming to practical usage of the hit-or-miss transform for template matching, call for a new definition able to achieve morphological template matching in real-life situations. We will now introduce such a new definition.

3.2. Towards a more intuitive definition

In multivariate images, template matching can benefit from user knowledge both on spatial and spectral point of views. The main difficulty faced by template matching operators (such as the morphological hit-or-miss transform) is how to combine these two kinds of information. While the spatial information such as the shape and the size of the sought object can be easily provided by a user, its combination with spectral information is not trivial. Moreover, when dealing with complex patterns, defining a single pair of structuring elements or even functions may be very challenging for the user.

Thus we propose to consider another strategy to design the structuring elements. When seeking for a predefined complex template, we believe that the user is able to describe this template by a set of elementary units. Each of these units describes a particular feature of the template, combining some spatial and spectral information. More precisely, it consists in an expected spectral response in some spatial area. To represent such a knowledge, we define each particular feature by an extended structuring element as follows. The spatial properties (shape and size of the area where spectral knowledge is known) are provided by the structuring element similarly to existing hit-or-miss definitions.

For each area, the spectral information consists of an expected intensity or value in a given spectral band. Thus it can be either lower or higher bounded by a predefined threshold. This results in several spectral properties defined for each structuring element: the spectral band it is related to, the kind of threshold used (either low or high threshold) and the threshold value.

Compared to the classic definition of the hit-or-miss transform, we consider here a set of extended structuring elements (not necessarily only 2) to be involved in the matching process. While our proposition manages spatial information similarly to previous approaches, we attribute here a particular attention to the spectral information. Indeed, contrary to existing multivariate strategies where the structuring element is shared between all spectral bands (in the marginal strategy) or is intrinsically multivariate (in the vectorial strategy), the proposed extended structuring element is dedicated to a single spectral band. By this way, the user can more easily design his set of structuring elements based on his prior knowledge on the sought template. The use of low and high thresholds helps to ensure the robustness of the template matching process, provides a more practical and realistic way to formulate prior spectral knowledge (compared to contrast-based definitions) and may be seen somehow as a generalisation of the initial hit-or-miss transform (where the notions of high and low are intrinsically brought by the use of foreground and background structuring elements, or erosion and dilation operators).

For a given extended structuring element k from the set K, we then write F_k its spatial pattern (combining shape and size information), b_k its spectral band, t_k its threshold or bound and ϕ_k its related operator (which can be either dilation δ or erosion ε, corresponding respectively to a high or low threshold). The spatial pattern is not expected to be constant over the different image bands. Let us note that several extended structuring elements may consider the same spectral band, under the assumption that they are consistent together (e.g. do not ask the signal to be both lower and higher than the same threshold !). Conversely, some image band might be of no interest for a given template matching task and thus not be related to any SE. The fitting step consists in checking, for each analysed pixel, if its neighbourhood matches the set of extended structuring elements. A pixel will be matched if and only if its neighbourhood fits all the structuring elements, i.e. the following condition holds:

$$\mathrm{MHMT}_K(\mathbf{w})(\mathbf{p}) \text{ fits iff } \forall\, k \in K, \quad \begin{cases} \varepsilon_{F_k}(w_{b_k})(\mathbf{p}) \geq t_k, & \text{if } \phi_k = \varepsilon \\ \delta_{F_k}(w_{b_k})(\mathbf{p}) < t_k, & \text{otherwise} \end{cases} \tag{13}$$

Other fusion options are available to merge the individual fitting results, but the conjunction is of course to be preferred since it ensures that the proposed operator possesses a consistent behavior with common morphological transforms (but this fusion operator may be replaced with some others to improve robustness, as it will be discussed further in this paper). Similarly to existing definitions, the fitting is followed by a valuation step which aims at giving a resulting value to all matched pixels (we assume that unmatched pixels are set to $\perp$). But contrary to previous works, here we do not deal with a single pair of foreground/background (or erosion/dilation) structuring elements, but rather have to deal with a whole set of extended structuring elements with various properties (shape and size but also spectral band and threshold value, as well as the threshold or operator type). Since our aim is to measure how well a pixel (and its neighbourhood) fits a complete set of extended structuring elements, we propose to first perform a valuation for each individual structuring element. In order to measure the quality of each individual fitting procedure, we can rely on the erosion or dilation result and the considered threshold, instead of both erosion and dilation as in existing hit-or-miss definitions. Thus we propose to compute the difference between the morphologically processed (either eroded or dilated) pixel and the (either low or high) threshold, i.e.

$$\mathrm{MHMT}_k(\mathbf{w})(\mathbf{p}) = \begin{cases} \varepsilon_{F_k}(w_{b_k})(\mathbf{p}) - t_k, & \text{if } \phi_k = \varepsilon \\ t_k - \delta_{F_k}(w_{b_k})(\mathbf{p}), & \text{otherwise} \end{cases} \tag{14}$$

thus ensuring a strict positive result for each matched pixel. However, we cannot assume that in practice multivariate images will always contain comparable spectral bands. In other words, the different spectral components of a multivariate image may not share the same value ranges. Thus, we propose to introduce a normalisation step, resulting in a new definition for the individual valuation step:

$$
\begin{aligned}
\mathrm{MHMT}_k(\mathbf{w})(\mathbf{p}) &= \begin{cases} \left(\varepsilon_{F_k}(w_{b_k})\,(\mathbf{p}) - t_k\right) / \left(w_{b_k}^{+} - t_k\right), & \text{if } \phi_k = \varepsilon \\ \left(t_k - \delta_{F_k}(w_{b_k})(\mathbf{p})\right) / \left(t_k - w_{b_k}^{-}\right), & \text{otherwise} \end{cases} \\
&= \begin{cases} \left(\varepsilon_{F_k}(w_{b_k})\,(\mathbf{p}) - t_k\right) / \left(w_{b_k}^{+} - t_k\right), & \text{if } \phi_k = \varepsilon \\ \left(\delta_{F_k}(w_{b_k})(\mathbf{p}) - t_k\right) / \left(w_{b_k}^{-} - t_k\right), & \text{otherwise} \end{cases}
\end{aligned} \tag{15}
$$

with $[w_i^-, w_i^+]$ the predefined value range of the spectral band w_i (and of course the assumptions $t_k \neq w_{b_k}^-$ and $t_k \neq w_{b_k}^+$). The normalisation is achieved by $\left(w_{b_k}^+ - t_k\right)$ or $\left(w_{b_k}^- - t_k\right)$ in order to obtain a valuation in [0;1]. Once the individual valuations have been computed, it is then necessary to assign a unique value to each matched pixel. As discussed in section 2.3, there is no unique valuation scheme for multivariate hit-or-miss. Indeed, we can even keep the set of individual valuations as the final result if we are interested in the quality of the fit for each individual pattern. But usually, a single scalar value is expected and we propose to build it by relying on a fusion rule. To ensure coherence with the previous fitting step (where a conjunction rule has been used to merge individual fitting results), we propose here to rely on a T-norm, e.g. either the product or the minimum, leading respectively to the following definitions:

$$
\begin{aligned}
\mathrm{MHMT}_K^{\mathrm{prod}}(\mathbf{w})(\mathbf{p}) &= \begin{cases} \prod_{k \in K} \left(\mathrm{MHMT}_k(\mathbf{w})(\mathbf{p})\right) & \text{if } \forall\, k \in K,\ \mathrm{MHMT}_k(\mathbf{w})(\mathbf{p}) > 0 \\ 0 & \text{otherwise} \end{cases} \\
&= \prod_{k \in K} \Big(\max\left(\mathrm{MHMT}_k(\mathbf{w})(\mathbf{p}),\ 0\right)\Big)
\end{aligned} \tag{16}
$$

$$
\begin{aligned}
\mathrm{MHMT}_K^{\mathrm{min}}(\mathbf{w})(\mathbf{p}) &= \begin{cases} \min_{k \in K} \left(\mathrm{MHMT}_k(\mathbf{w})(\mathbf{p})\right) & \text{if } \forall\, k \in K,\ \mathrm{MHMT}_k(\mathbf{w})(\mathbf{p}) > 0 \\ 0 & \text{otherwise} \end{cases} \\
&= \min_{k \in K} \Big(\max\left(\mathrm{MHMT}_k(\mathbf{w})(\mathbf{p})\right),\ 0\Big)
\end{aligned} \tag{17}
$$

3.3. Definitions within the morphological framework

In the previous definition, we have shown in a comprehensive way how easy it could be to define template matching parameters (i.e. structuring elements) based on some prior spatial and spectral knowledge. Let us note that the proposed hit-or-miss transform can also be defined within the scope of mathematical morphology and we show here that it shares some common properties with other definitions related to binary, grayscale and multivariate cases.

3.3.1. Link with binary morphology

In the previous section, we have introduced within our HMT definition the use of thresholds as lower or higher bounds of user-provided spectral knowledge for some spatial areas of the sought pattern. The thresholding process can also be considered as a pre-processing step to be applied before the hit-or-miss transform. In such a case, the thresholded binary images become then compatible with binary mathematical morphology and we can study the link between our proposed operator and binary HMT. To do so, for each extended structuring element k, we define a binary thresholded image X_k from the input multivariate image $\mathbf{w}$ as follows:

$$X_k = \{\mathbf{p} \mid w_{b_k}(\mathbf{p}) \geq t_k\} \tag{18}$$

The fitting step of our proposed MHMT defined in Eq. (13) can then be written

$$\mathrm{MHMT}_K(\mathbf{w})(\mathbf{p}) \text{ fits iff } \forall\, k \in K, \quad \mathbf{p} \in \mathrm{HMT}(X_k) \tag{19}$$

where individual binary hit-or-miss results are given by

$$\mathrm{HMT}(X_k) = \begin{cases} \varepsilon_{F_k}(X_k) & \text{if } \phi_k = \varepsilon \\ (\delta_{F_k}(X_k))^c & \text{otherwise} \end{cases} \tag{20}$$

$$= \begin{cases} \varepsilon_{F_k}(X_k) & \text{if } \phi_k = \varepsilon \\ \varepsilon_{F_k}(X_k^c) & \text{otherwise} \end{cases} \tag{21}$$

As we can see, this formulation is coherent with the binary case given in Eq. (2). Moreover, the binary HMT can simplify into a single erosion (i.e. $\mathrm{HMT}(X'_k) = \varepsilon_{F_k}(X'_k)$) if we take early care of the operator type ϕ_k during the thresholding process:

$$X'_k = \{\mathbf{p} \mid w_{b_k}(\mathbf{p}) \geq t_k \text{ if } \phi_k = \varepsilon,\ \ w_{b_k}(\mathbf{p}) < t_k \text{ if } \phi_k = \delta\} \tag{22}$$

The fitting step of our proposed MHMT can then be seen as a binary HMT applied on some thresholded images, thus making possible to introduce spectral lower and higher bounds in the template matching process. Due to the multivariate input, the valuation step of our operator is more powerful than the binary HMT one which only returns 0 for unmatched pixels and 1 for matched pixels. Nevertheless, this binary behavior can be achieved by our operator, by applying a thresholding step on the final result given by Eq. (16) or (17):

$$\mathrm{MHMT}_K^{\text{binary}}(\mathbf{w})(\mathbf{p}) = \begin{cases} 1 & \text{if } \forall\ k \in K,\ \mathrm{MHMT}_k(\mathbf{w})(\mathbf{p}) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

3.3.2. Link with greyscale/marginal morphology

We have shown that the lower and higher spectral bounds involved in our MHMT definition can be expressed as a preprocessing thresholding step before using binary HMT. These bounds can also lead to the definition of structuring functions, to be used within the scope of greyscale HMT definitions given in Sec. 2.2. Since we are dealing with multivariate images, we rather consider here multivariate structuring functions to be used following a marginal strategy (the case of vectorial strategy will be discussed in Sec. 3.3.3).

More precisely, let us focus on the HMT definition from Soille given in Eq. (6). Extension of this greyscale operator can be performed following the marginal strategy given in Eq. (10). To avoid the same spatial processing of all spectral bands during the template matching process, the pair of structuring elements (F, G) may be replaced with a pair of multivariate structuring functions $(\mathbf{f}, \mathbf{g})$, thus resulting in the following definition:

$$\mathrm{HMT}_{\mathbf{f},\mathbf{g}}^{\text{marginal Soille}}(\mathbf{w}) = \left(\mathrm{HMT}_{f_1,g_1}^{\text{Soille}}(w_1),\ \ldots,\ \mathrm{HMT}_{f_n,g_n}^{\text{Soille}}(w_n)\right) \tag{24}$$

where

$$
\mathrm{HMT}^{\mathrm{Soille}}_{f_i,g_i}(w_i)(\mathbf{p}) = \begin{cases} \varepsilon_{f_i}(w_i)(\mathbf{p}) - \delta_{g_i^*}(w_i)(\mathbf{p}) & \text{if } \varepsilon_{f_i}(w_i)(\mathbf{p}) \geq \delta_{g_i^*}(w_i)(\mathbf{p}) \\ 0 & \text{otherwise} \end{cases}
$$

$$
= \max\{\varepsilon_{f_i}(w_i)(\mathbf{p}) - \delta_{g_i^*}(w_i)(\mathbf{p}), 0\} \tag{25}
$$

In our definition, we have rather considered a set K of extended structuring elements, each being defined by shape F_k, a spectral band b_k, a bound value or threshold t_k and a bound type (lower/higher) or operator ϕ_k. Nevertheless, these extended structuring elements may be used all together to build as follows the two multivariate structuring functions $\mathbf{f}$ and $\mathbf{g}$ required in Eqs. (24) and (25):

$$
f_i(\mathbf{p}) = \max_{k \in K} (t_k \mid p \in F_k,\ b_k = i,\ \phi_k = \varepsilon) \tag{26}
$$

$$
g_i(\mathbf{p}) = \min_{k \in K} (t_k \mid p \in F_k,\ b_k = i,\ \phi_k = \delta) \tag{27}
$$

thus we consider for each spectral band, the most restrictive conditions (i.e. the highest bound for erosion, the lowest bound for dilation) available in each location of the template. Let us observe that this new formulation is equivalent to the set of extended structuring elements given in Sec. 3.2. The fitting conditions given in Eq. (13) can then be rewritten using $\mathbf{f}$ and $\mathbf{g}$:

$$
\mathrm{MHMT}_{\mathbf{f},\mathbf{g}}(\mathbf{w})(\mathbf{p}) \text{ fits iff } \forall\, b \in B,\ \ \varepsilon_{f_b}(w_b)(\mathbf{p}) \geq 0 \text{ and } \delta_{g_b}(w_b)(\mathbf{p}) < 0 \tag{28}
$$

which is a particular (i.e. more restrictive) case of the fitting condition given in Eq. (25), since we require here a positive erosion result, a negative dilation result and that the fitting occurs in all spectral bands addressed by the template or set K.

3.3.3. Link with vectorial morphology

In this section, we study the links between our proposition and current multivariate HMT approaches, both with the marginal and vectorial strategies.

To ensure a purely marginal fitting (i.e. when a pixel is matched or not in a given spectral band only relying on values of its neighbourhood in this band), we should replace the condition given in Eq. (13) by

$$\mathrm{MHMT}_K(w_i)(\mathbf{p}) \text{ fits iff } \forall\, k \in K,\ b_k = i \quad \begin{cases} \varepsilon_{F_k}(w_{b_k})(\mathbf{p}) \geq t_k, & \text{if } \phi_k = \varepsilon \\ \delta_{F_k}(w_{b_k})(\mathbf{p}) < t_k, & \text{otherwise} \end{cases} \tag{29}$$

which differs from Eq. (25) only by the use of additional thresholds set intermediately between erosion and dilation.

Moreover, we can also replace valuation procedures we suggested in Eqs. (16) and (17) by

$$\mathrm{MHMT}_K(w_i)(\mathbf{p}) = \begin{cases} \min\limits_{k\in K}\left(\mathrm{MHMT}_k(w_i)(\mathbf{p}) \mid b_k = i,\ \phi_k = \varepsilon\right) & \\ + \min\limits_{k\in K}\left(\mathrm{MHMT}_k(w_i)(\mathbf{p}) \mid b_k = i,\ \phi_k = \delta\right) & \text{if } \mathrm{MHMT}_K(\mathbf{w}) \text{ fits} \\ 0 & \text{otherwise} \end{cases} \tag{30}$$

thus resulting in the same multivariate valuation as the one obtained by applying the Soille HMT in a marginal fashion with structuring functions, Eq. (25).

Since multivariate images may be processed with vectorial morphology [1], it is worth studying also the link with the vectorial HMT definition given by Aptoula et al. [2]. To be compatible with the vectorial HMT definition recalled in Eq. (11) (apart for the additional positive/negative constraint imposed on erosion/dilation), we need to express our fitting condition given in Eq. (28) with a vectorial ordering:

$$\varepsilon_{\mathbf{f}}(\mathbf{w})(\mathbf{p}) \geq_v \mathbf{0} >_v \delta_{\mathbf{g}}(\mathbf{w})(\mathbf{p}) \tag{31}$$

where $\leq_v$ is simply defined as the componentwise (or marginal) ordering:

$$\forall\, \mathbf{v}, \mathbf{v}' \in \mathbf{T},\ \ \mathbf{v} \leq_v \mathbf{v}' \ \Leftrightarrow\ \forall\, i \in \{1, \ldots, n\},\ v_i \leq v'_i \tag{32}$$

which ensures the mandatory property of (vertical) translation invariance and offers a behavior coherent with Eq. (28) when used in combination with the null vector $\mathbf{0}$.

Let us observe that, contrary to the standard marginal definition given in Eq. (10), we are here not affected by the two drawbacks of the componentwise ordering. Indeed, there is no need to preserve input vectors since erosion and dilation multivariate results are not used directly, but either in a comparison for the fitting or in a scalar projection for the valuation. Moreover, even if correlated information among image channels is ignored during the computation of erosion and dilation, the final comparison with the null vector requires all image components to fulfill the given condition. It is thus a way to combine information spread on the different image channels.

In order to achieve the same behavior as vectorial extensions of Soille and Ronse operators given in Eq. (11), we may respectively apply the Euclidean norm on the valuation result given in Eq. (30) and consider the following valuation function:

$$\mathrm{MHMT}_K(w_i)(\mathbf{p}) = \begin{cases} \min_{k \in K} \left(\mathrm{MHMT}_k(w_i)(\mathbf{p}) \mid b_k = i,\ \phi_k = \varepsilon\right) & \text{if } \mathrm{MHMT}_K(\mathbf{w}) \text{ fits} \\ \perp & \text{otherwise} \end{cases} \tag{33}$$

but in this last case (and in this last case only), the ordering in use may be inadequate and provide false color and spectral signatures in the output image. Nevertheless, Ronse valuation is rather dedicated to image filtering than to template matching.

3.4. Comparison with related work

In this section, we review the main differences between our proposed MHMT and related works [2, 15, 18].

First, the MHMT parameters (i.e. the templates or structuring elements) are not expressed as a pair of multivariate structuring functions, but more easily as a set of elementary units composed of spatial areas where knowledge of the expected signal behavior (lower or higher bounded by a threshold) in a given spectral band is known. It also allows to easily deal with spectral bands which are not defined on the same value range.

Next, the fitting step of our proposition is more constraining than existing definitions which only requires that the erosion with the foreground structuring element should be greater than the dilation with the background structuring element. Indeed, we also impose that the former has to be positive while the latter is negative. By doing so, we lose the vertical translation invariance (i.e. $\mathrm{MHMT}(\mathbf{w}) \neq \mathrm{MHMT}(\mathbf{w_t})$ where $\mathbf{w_t}(\mathbf{p}) = \mathbf{w}(\mathbf{p}) + \mathbf{t}$) present in existing HMT definitions which are true contrast-based operators. But we are then able to gather the relative contrast information with some information of absolute nature, i.e. the minimum or maximum spectral values in some predefined spatial areas of the template.

Moreover, since the two existing approaches for extending the hit-or-miss transform to multivariate images present some drawbacks (the marginal strategy ignore correlated information while the vectorial strategy needs to select the appropriate vectorial ordering), we suggested an intermediate approach to avoid these drawbacks. Applying the fitting individually on each spectral band (or more precisely, with each extended structuring element) prevents us from selecting a vectorial ordering, while merging these individual fitting steps in a global one enable us (in some way) to take care of the correlated information spread over the different image channels.

As far as the valuation is concerned, we propose several models which are able to exploit individual fitting results. Moreover, these procedures may be applied within a normalisation framework, thus making possible to deal with multivariate images with heterogeneous spectral bands and to produce results compatible with inputs of fuzzy or soft logic.

Finally, as expected, our MHMT can simplify in existing definitions when applied on a single greylevel or binary images. Let us also observe that even in the multivariate case, our operator will lead to a greyscale result, which can be easily processed by the end-user (e.g. display, thresholding, etc). As we will see next, the proposed definition also allows an efficient implementation.

3.5. Implementation

We have previously given several formulations for the proposed multivariate hit-or-miss transform. The intuitive definition, despite being less theoretically sound, leads to various optimisation schemes which will be presented here. Together they lead to an efficient implementation of the HMT transform, in particular for the fitting step (which is in fact the most important part of the template matching process).

Contrary to a multivariate structuring element or function processed globally as proposed in [2], here we can decompose the set of structuring elements into elementary patterns and process them sequentially. Since the global fitting process requires a pixel to fit all the structuring elements to be kept, we can stop the process as soon as one of the SE is not matched by the HMT. Thus we perform only an incomplete processing of the set of SE.

Moreover, since within the HMT process the result of the erosion (or dilation) is compared to a fixed threshold, we can avoid the minimum (or maximum) computation which has to be performed before the threshold comparison. Indeed, each pixel can be directly compared to the given threshold: as soon as a pixel does not fulfill the threshold condition, the processing can be stopped since the SE will not be matched. Using this formulation, one does not need to determine the minimum or maximum anymore. This optimisation leads to an incomplete processing of each SE.

The fitting algorithm, based on these optimisations, is provided in Algorithm 1. Let us note that it is even more efficient if the set of SEs is sorted with the most discriminative SEs first. By this way, the fitting process can stop sooner with unmatched pixels and global computation time is limited.

4. A practical guide for MHMT-based template matching

The proposed MHMT operator aims to extract features using spatial and spectral patterns. It can then be used to build template matching process based on mathematical morphology. In this section, we present how to design such

Data: Im: Image; SE: Array of SEs
Result: FIm : Fitting image

```
initialisation;
for i:each pixel of Im do
    for j:each structuring element of SE do
        if φ_SE[j] = δ then
            for k:each pixel of F_SE[j] do
                if Im_b_SE[j](i + k) ≥ t_SE[j] then
                    FIm(i) = false;
                    GOTO next pixel of Im;
                end
            end
        end
        if φ_SE[j] = ε then
            for k:each pixel of F_SE[j] do
                if Im_b_SE[j](i + k) < t_SE[j] then
                    FIm(i) = false;
                    GOTO next pixel of Im;
                end
            end
        end
        FIm(i) = true;
    end
end
Result= FIm;
```

Algorithm 1: Improved MHMT fitting

a system and compare it with classical template matching methods. We then propose several application cases related to the field of remote sensing, before discussing possible improvements of our method.

4.1. Design morphological template matching for different features

To be successfully applied, morphological template matching required to be adapted to the specificities of the desired pattern. In this section, we distinguish two types of patterns (discontinuity and shape) and explain how to design dedicated template-matching systems using the proposed MHMT operator. Indeed, each feature will require the system to ensure some properties (e.g. robustness against orientation or scale), thus impacting the design of the template matching system.

4.1.1. Discontinuity

Discontinuity is an abstract feature: indeed, it is not considered as a straight visual feature but rather denotes the limit between two specific areas. Discontinuity extraction is used, for example, by coastline delineation methods. In such cases, we assume it exists two sets of opposite SEs, each of them defining one given area. A MMTM method dedicated to particular discontinuity extraction is built in two steps:

- The two areas have to be defined spectrally and spatially, but the spatial definition is only related to the depth or width of the feature.
- Apart for the detection of a specific directional discontinuity (e.g. only vertical discontinuity), there is a need of applying MHMT at different orientations of the SEs.

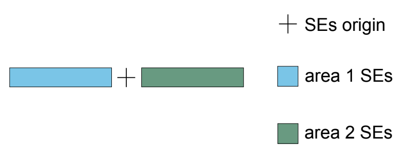


Figure 4: Example of SEs for an oriented discontinuity extraction

We notice that the two sets of SEs do not have to be connected, the proposed MHMT operator can handle an uncertainty area (similarly to the standard HMT definition and usage).

4.1.2. Object

Object may be seen as the classical feature of template matching. Specific object extraction is used to find a specific object in an picture. Here we assume there are two sets of SEs: one is defining the object (being possibly as small as a single point) while the other is defining what is not the object (e.g. , the background). A MMTM method dedicated to object extraction is built in three steps:

- The SEs representing the object are defined both in terms of spatial and spectral information
- The SEs representing what is not the object are defined both in terms of spatial and spectral information
- Depending of the object (fixed size or not, different orientations or not), the MHMT can be applied with SEs at different orientations and/or different scales.

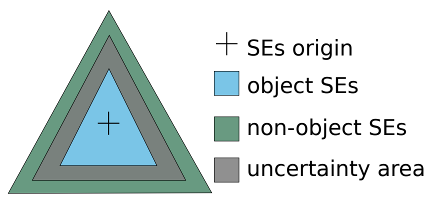


Figure 5: Example of SEs for object extraction

In order to be able to match the object even with discretisation problems (e.g., stairing effect), we advise letting an uncertainty area between the different sets of SEs.

4.1.3. Robustness

To be really useful in practice, the template matching operator needs to ensure a certain level of robustness. Here we discuss the robustness of the MHMT against orientation, scale, translation and contrast.

The robustness against orientation and scale can be ensured by applying the template in different orientations with different sizes as it is used in the next section. This is possible thanks to the hard fitting step, which prevents from combining different fitting values. Indeed, a template is matched if it fits for at least one configuration (size, orientation). The computation of the valuation is the following: if the template is matched for different orientations/sizes, the resulting valuation is computed as the highest of the single valuations since it is expected to represent the best matching of the template. While our operator is applied on each pixel of the image, it is of course robust to translation. But the MHMT is not robust to contrast change. Indeed, since the expert set the threshold values, a change of contrast may lead to a pixel value modification which will be high enough to break the threshold conditions.

4.2. Setting the SE parameters

The settings of the SE parameters (b, t, ϕ) is expected to be made by an expert based on her domain knowledge. In our context (i.e., remote sensing), the expert knows (at least partially) the spectral signature of the desired object and of its environment, or of the areas surrounding the desired discontinuity. This spectral information depends on the type of sensor and eventually external factors (e.g., season for remote sensing imagery). Obviously, the goal here is not to use all possible constraints, since it would be computationally expensive. We rather consider only the most discriminant spectral information of the object under study when compared to its environment. For instance, if we consider spectral knowledge given in Table 1, we can observe an important overlap of spectral signatures of sea and land in bands *Green* and *Red*. Thus, the expert will most probably not use these spectral bands as discriminative information in the template matching process. Setting parameters w.r.t. bands *Blue* and *Near-Infrared* looks much more relevant since the related information is more reliable to distinguish between sea and land.

To ease the SE parameter setting step, intuitive and interactive approaches may be considered. For instance, we consider here the case of drawing scribbles on objects of interest which is a very common method for interaction [6, 22]. It then allows to learn the spectral range of the desired features and to automatically set the SE parameters in order to achieve the highest discriminative power of the template matching method. This principle is illustrated in Fig. 6 where the user has drawn markers on two warehouses and on the background:

Band	**Sea**			**Land**		
Blue	173	–	193	134	–	165
Green	215	–	269	147	–	241
Red	119	–	167	66	–	177
Near Infrared	55	–	84	126	–	575

Table 1: Spectral ranges from sea and land obtained from a QuickBird image.

from this user input, the system is able to obtain the spectral range of the templates defined from the scribbles (Tab. 2). The final step is then performed either by the user who analyse these spectral ranges to select appropriate bands and thresholds, or directly by a machine learning algorithm to identify the best decision functions.

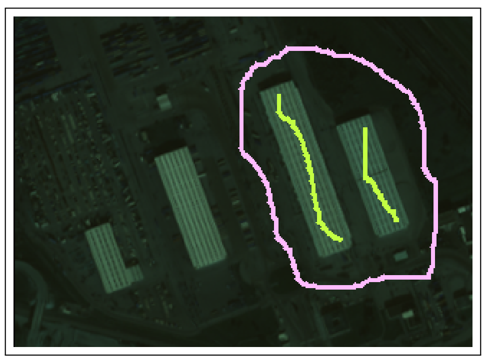


Figure 6: Scribbles drawn on some warehouses and background to learn SE parameters from a QuickBird Image (©Digitalglobe).

4.3. Comparison with classical TM methods

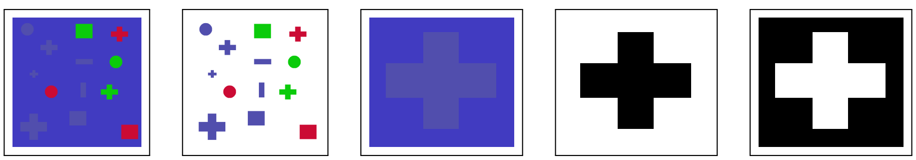
In this section, we compare the MHMT with three classical template matching methods: a purely spectral method, the multi-thresholding approach; a purely spatial method, the vectorial HMT [2] with flat structuring elements; and a mixed method, the correlation technique [8].

Figure 7 presents the 24 bits RGB image and the objects it contains which

Band	Warehouses	Background
Blue	336 - 556	142 - 323
Green	511 - 857	161 - 476
Red	381 - 635	82 - 344
Near Infrared	362 - 612	93 - 431

Table 2: Spectral range of the scribbles drawn by the user on figure 6.

are crosses, rectangles, circles and squares of different colors. This figure also shows the desired template, a cross with a color close to the background color and the shape of the two structuring elements MHMT and VHMT will be using.



a b c d e

Figure 7: a) Original image; b) Image without background; c) Template to detect (magnified); d) Foreground structuring element shape (magnified); e) Background structuring element shape (magnified)

Multi-thresholding is applied here with a double thresholding of each original image band. Values used by the double thresholding are the values surrounding template values. Vectorial HMT is applied using the two flat structuring elements shown in figure 7 and lexicographical ordering. Correlation used the whole template. MHMT is applied using SEs defined as follows:

$$\Omega = \begin{array}{lllll} \{F_{s^1} = & Foreground\ SE, & b_{s^1} = \mathrm{RED}, & t_{s^1} = 80, & \phi_{s^1} = \varepsilon\} \\ \{F_{s^2} = & Foreground\ SE, & b_{s^2} = \mathrm{GREEN}, & t_{s^2} = 77, & \phi_{s^2} = \varepsilon\} \\ \{F_{s^3} = & Background\ SE, & b_{s^3} = \mathrm{BLUE}, & t_{s^3} = 192, & \phi_{s^3} = \varepsilon\} \end{array} \tag{34}$$

Visual results of each method presented in figure 8 show that only the correlation and the MHMT have succeeded in detecting the specified cross template.

Nevertheless, the correlation requires a subsequent thresholding to obtain a binary decision. Moreover, as shown in table 3, the MHMT achieves a more efficient processing than the correlation. Let us notice however that we are using here the standard definition of the correlation operator, without any optimisation. We have not considered efficient algorithms working in the frequency space built from FFT since our method only deals with the spatial domain. Nevertheless, we could have relied on optimized spatial correlations methods [7] which would have most probably lead to similar computation time as our method, but such approach are approximations of correlation and a deep comparison on this issue was beyond the scope of this paper. Only the multi-thresholding is faster than the MHMT, but since it ignores spatial information, it extracts a lot of false positive objects. VHMT gives only one false positive but is the slowest of the four methods. Computation times are obtained on Core i7 Q720 CPU using a Java implementation.

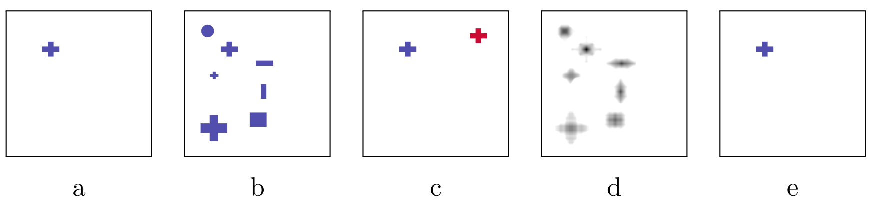


Figure 8: a) MHMT after reconstruction; b) Multi-thresholding; c) VHMT with lexicographical ordering after reconstruction; d) Correlation; e) Correlation after thresholding and reconstruction

In conclusion, compared to these standard template matching methods, the MHMT provides relevant results while ensuring a low computation time. These abilities will be further studied by some real applicative scenarios.

4.4. Applications

To further show the relevance of our new operator, we propose here a couple of experiments related to the remote sensing domain. Images considered here are 4-band QuickBird images with a spatial resolution of 2.4 meters per pixel.

Method	**Computing Time** (sec)	**False Positive**
MHMT + rec.	0.3	0
Correlation + thresh. + rec.	167.0	0
Vectorial + rec.	224.1	1
Multi-thresholding	0.1	6

Table 3: Comparison of the different methods in terms of computing time and false positives

Spectral bands are related to near infra-red, red, green and blue wavelengths. Image size is equal to 1780×782 pixels in the first experiment (Fig. 10) and 627×376 pixels in the second experiment (Fig. 13). Let us also note that pixel values are coded over 11 bits (2,048 possible values), which has no influence on our method (apart from the values t_s).

4.4.1. Border extraction in satellite VHR images

As a first application example in remote sensing, we consider the case of satellite images with a very high spatial resolution (VHR). Template matching on such images can focus on some predefined borders, coastline being one of the most representative examples when dealing with coastal remote sensing.

Here, we are in a context of discontinuity extraction as coastline is typically defined as the border between sea and land. Its automatic extraction from digital image processing is a very topical issue in remote sensing imagery [5]. Even if some methods have been proposed for low or medium spatial resolution, none are relevant on very high spatial resolution (VHR) satellite imagery where a pixel represents an area lower than $5 \times 5\ m^2$.

Since a coastline is well defined both with spatial and spectral information, the proposed MHMT can be seen as a relevant tool to perform its extraction. A basic assumption would be that only two SEs are necessary to identify the two parts around the border (i.e. sea and land) as shown by S^1 and S^2 in Figure 9. However, to be able to distinguish between coastlines and other water-land borders (e.g. lake border, river bank, etc), we involve an additionnal SE to represent deep sea (or at least water further from the coastline) as illustrated by S^3 in Figure 9.

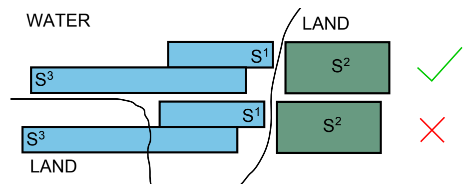


Figure 9: Spatial definition of SEs used for coastline extraction with matching and unmatching conditions

In order to limit the computation time, the shape of each SE can be limited to a single line segment (either on the sea or the land part of the border). However with this configuration, it would be only possible to detect coastline for a single orientation. To ensure rotation-invariant template matching, multiple orientations of the SEs are necessary. To do so and to be able to consider all possible coastline orientations, the application of the MHMT is performed in various directions. The interested reader can find more details in [14].

Relevance of MHMT for the extraction of coastline in VHR imagery can be observed in Fig. 10. In order to demonstrate the effectiveness of the MHMT-based approach, we compare it to three classical coastline extraction methods: Bagli and Soille [3], Heene and Gautama [9] and Jishuang and Chao [10]. We use two evaluation measures which respectively characterize the accuracy and the precision of the methods. Accuracy is evaluated through *mean distance*, measuring the average distance between the extracted coastline and the reference coastline (ground truth). It is computed by averaging distances from all extracted pixels to the closest pixels of the reference. Precision is evaluated by measuring the *number of false positive connected components*. Indeed, it penalizes methods which may produce a very accurate result (i.e. a coastline equal to the reference) but also many outliers or irrelevant (coastline) regions. To do so, we first perform a connected component (CC) labeling and then enumerate all

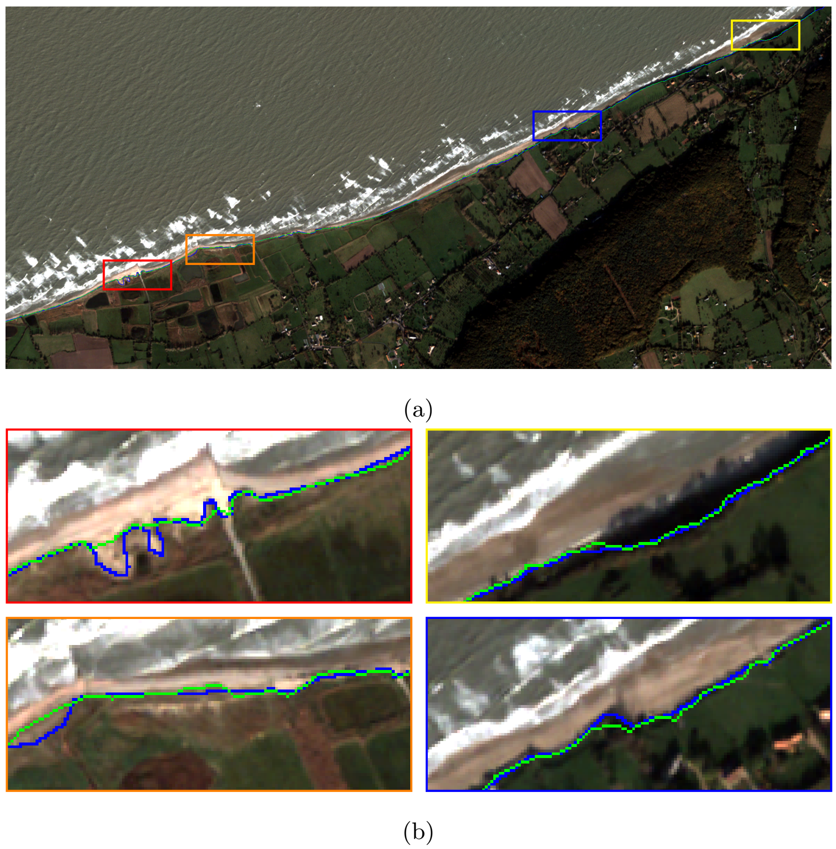


Figure 10: (a) Coastline extraction on Normandy coast from a QuickBird image (©Digitalglobe), extracted coastline in green and reference coastline in blue, (b) details of the extracted coastline.

CC which do not overlap the coastline reference. Results of the comparison are given in Table 4 and show that MHMT-based method is closer to the reference coastline than other methods and give no false positives, contrary to two of the compared methods (visual results are shown in figure 11).

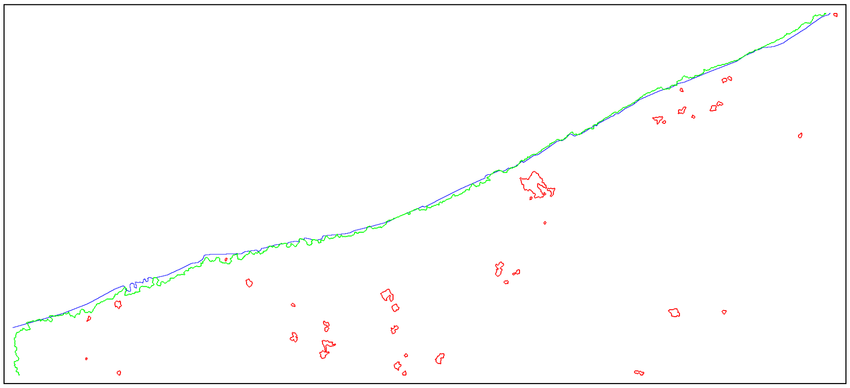
Method	Mean Distance (per pixel)	False Positive (in terms of CC)
MHMT-based	1.00	0
Bagli and Soille [3]	6.90	35
Heene and Gautama [9]	2.52	30
Jishuang and Chao [10]	2.10	0

Table 4: Error measures produced by the different coastal extraction methods

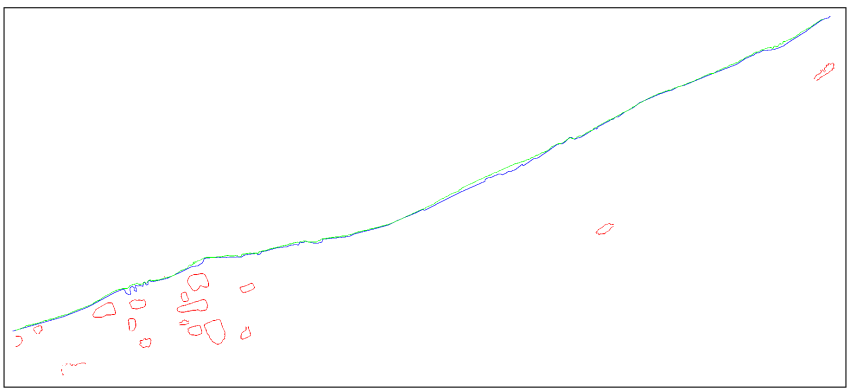
4.4.2. Petroleum tank extraction in satellite VHR images

The second application deals with object extraction and more precisely petroleum tank extraction. We consider a study site located in the harbor of Le Havre in France. Petroleum tanks are cylindric like the other tanks, but in this harbor, they are white in order to distinguish them from other tanks (and thus look brighter in the image). So, to solve the problem of petroleum tank extraction, spatial information (cylindric shape) is not enough, we need to use the spectral information (petroleum tanks are white). This is the reason why we use MHMT to extract these objects. MHMT is used with two sets of SEs, one defining the petroleum tanks, the other defining what is around the petroleum tanks. Their shapes are shown in figure 12.

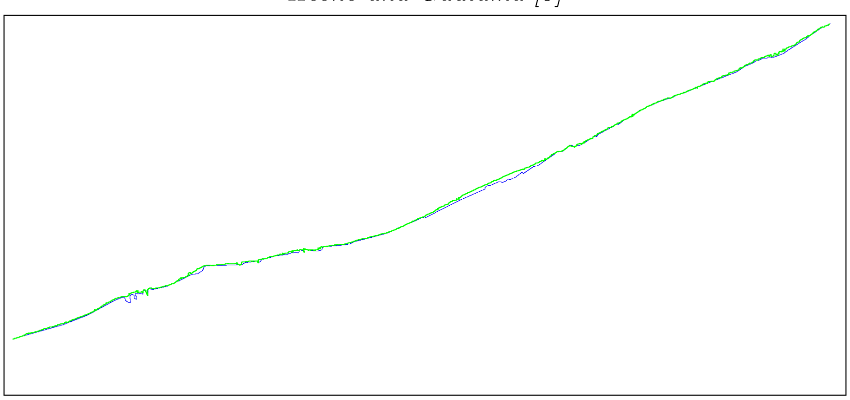
Since there is not a unique size for petroleum tanks, MHMT has to be applied with various SE sizes. The spectral parameters are defined by an expert and given in equation 35. The spatial parameters are the following: we consider a circular SE of radius r varying from 2 to 20 pixels, while the surrounding ring



Bagli and Soille [3]



Heene and Gautama [9]



Jishuang and Chao [10]

Figure 11: Coastline extraction by existing methods (extracted coastline in green, reference coastline in blue and false positive CC in red).

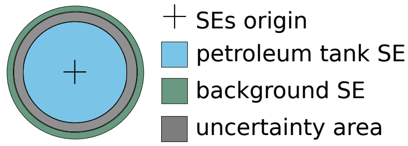


Figure 12: Spatial definition of SEs used for petroleum tank extraction

SE radius is a little bit larger (i.e. equal to $r+2$).

$$\Omega = \begin{array}{lllll} \{F_{s^1} = & circle & b_{s^1} = \text{BLUE}, & t_{s^1} = 450, & \phi_{s^1} = \varepsilon\} \\ \{F_{s^2} = & circle & b_{s^2} = \text{GREEN}, & t_{s^2} = 450, & \phi_{s^2} = \varepsilon\} \\ \{F_{s^3} = & circle & b_{s^3} = \text{RED}, & t_{s^3} = 450, & \phi_{s^3} = \varepsilon\} \\ \{F_{s^4} = & circle & b_{s^4} = \text{NIR}, & t_{s^4} = 450, & \phi_{s^4} = \varepsilon\} \\ \{F_{s^5} = & ring & b_{s^5} = \text{GREEN}, & t_{s^5} = 716, & \phi_{s^5} = \delta\} \end{array} \tag{35}$$

Shape *circle* represents object set of SEs while *ring* represents non-object set of SEs. NIR means Near Infra-Red band. In this example, thresholds have been easily set by the expert from a manual study of the multispectral image histogram. Indeed, tanks might be distinguished from the other objects and the image background based on their spectral signature. We rely here on peak values in the histogram to set adequate threshold values.

Applying MHMT with these parameters gives promising results as shown in figure 13. This satellite image contains 33 petroleum tanks, 32 were extracted, one was missed due to the presence of dark colors on its roof. There is also one false positive, a piece of a non petroleum tank was extracted as a petroleum tank but no entire other tank was extracted. MHMT appears here as an efficient solution to deal with the problem of specific tank extraction. Let us note that the problem of the missed tank can probably be solved by introducing some fuzziness in the MHMT operator.

4.5. Discussion

We have previously illustrated the potential of our proposed MHMT by two applications in remote sensing. To increase the relevance of MHMT in real-life

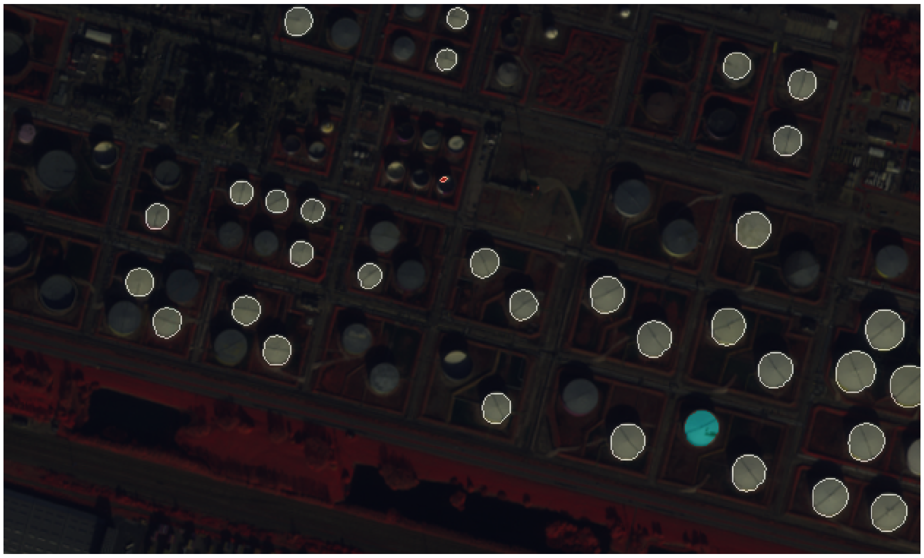


Figure 13: Petroleum tank extraction on Le Havre harbor QuickBird image(©Digitalglobe). Correct detections are surrounded by white boundary, false positive are given in red while false negative are given in cyan.

template matching problems, we believe some other improvements might be brought.

The first one is related to automatically learning the set of structuring elements. To replace expert knowledge, it might be useful to get the structuring elements through a machine learning strategy. Indeed, defining the structuring elements is not always straightforward. While the expert has a precise knowledge of the objects and the problem under study, it is not always easy for her to state this knowledge as a set of extended structuring elements. By providing some visual examples, machine learning strategy might be involved to determine the adequate set of extended SEs. In Sec. 4.2, we propose an intuitive scheme where the user is invited to draw scribbles on some objects of interest. From these scribbles, the different SE parameters can be learnt, depending on the type of desired feature (discontinuity or object), which can itself even be deduced from shape of the scribbles. In the discontinuity case, spectral parameters may be selected as the most specific spectral ranges between the two sides of the scribble. In the object case, they are rather determined from the most discriminant spectral ranges of the scribble pixels compared to their surrounding. Shape definition is more complex and definitely requires an in-depth study (out of the scope of this paper) to exploit state-of-the-art machine learning methods. It is also possible to perform parameter settings based on the usage of ontologies. Indeed, if a user has a well-defined ontology or knowledge base containing the objects of interest, most discriminant SE parameters might be easily extracted. Let us note that ontology-based solutions are also useful for automatically defining the spatial parameters (shape, size).

A second way of improvement is related to the hard behavior of the MHMT. Indeed, the proposed operator contains a fitting step which looks very sensible to noise due to its binary response. Let us observe however that it is a well-known drawback of most HMT operators. A HMT robust to noise has been proposed in Perret et al. [13], using a fuzzy definition of the HMT with standard structuring functions. This might be transposed to MHMT, in order to construct a fuzzy MHMT able to deal with noisy multivariate images without any additional cost regarding the SE definition. Bringing fuzziness to the MHMT would probably solve the problem encountered in section 4.4.2.

5. Conclusion and perspectives

Among the most common toolboxes for image analysis and processing, mathematical morphology offers a solution for the template matching problem with the operator called hit-or-miss transform (HMT). This operator considers two structuring elements or spatial templates to be matched within the input image, one for the foreground and one for the background. Despite a single definition on binary images, more complex images such as greyscale images have led to several propositions, in particular from Soille [18] and Ronse [15]. The case of multivariate images has been addressed only very recently [2], where requirements to extend existing greyscale HMT definitions to multivariate images are discussed. Such extensions require using a vectorial ordering, but selecting the adequate vectorial ordering in a given context may be tedious since a wide range of solutions are available [1]. A marginal strategy may also be involved, but it would result in the independent processing of all spectral bands, which is a major drawback when template matching relies on both spatial and spectral information.

So we address in this paper the problem of morphological template matching within multivariate images and introduce a new multivariate hit-or-miss transform. This operator does not need to define a vectorial ordering, but is still able to catch spectral information in the sought template. Moreover, its behavior can be easily understood, thus resulting in an intuitive way of parameter settings (i.e. template definition) and efficient implementations. Equipped with this operator, we discuss several template matching scenarios, considering discontinuity- or object-like templates. We finally demonstrated the relevance of our proposition by several template matching applications cases from the remote sensing field.

Further works will include a study of the operator's robustness against noise, which can be improved by using fuzzy structuring elements. Moreover, we also plan to facilitate the process of template designing by introducing a way to interactively learn the structuring elements to be involved in the template matching process.

6. Acknowledgements

The authors thank Anne Puissant from University of Strasbourg for the set-up and the evaluation of the coastline extraction experiment, Erchan Aptoula from Okan University for the English improvements and the ANR-JC ECOSGIL project for remote sensing data and financial support.

References

[1] Aptoula, E., Lefèvre, S., November 2007. A comparative study on multivariate mathematical morphology. Pattern Recognition 40 (11), 2914–2929.

[2] Aptoula, E., Lefèvre, S., Ronse, C., 2009. A hit-or-miss transform for multivariate images. Pattern Recognition Letters 30 (8), 760–764.

[3] Bagli, S., Soille, P., Oct. 2003. Morphological automatic extraction of paneuropean coastline from landsat etm+ images. In: Fifth International Symposium on GIS and Computer Cartography for Coastal Zone Management (COASTGIS).

[4] Barat, C., Ducottet, C., Jourlin, M., 2010. Virtual double-sided image probing: a unifying framework for non-linear grayscale pattern matching. Pattern Recognition 43, 3433–3447.

[5] Boak, E., Turner, I., 2005. Shoreline definition and detection: a review. Journal of Coastal Research 21 (4), 688–703.

[6] Boykov, Y. Y., Jolly, M. P., 2001. Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images. In: IEEE International Conference on Computer Vision. Vol. 1. pp. 105–112.

[7] Briechle, K., Hanebeck, U. D., 2001. Template matching using fast normalized cross correlation. In: SPIE Optical Pattern Recognition XII. Vol. 4387.

[8] Gonzalez, R. C., Woods, R. E., August 2002. Digital Image Processing (2nd Edition). Prentice Hall, p. 703.

[9] Heene, G., Gautama, S., 2000. Optimisation of a coastline extraction algorithm for object-oriented matching of multisensor satellite imagery. In: IEEE International Geoscience and Remote Sensing Symposium. Vol. 6. pp. 2632–2634.

[10] Jishuang, Q., Chao, W., 2002. A multi-threshold based morphological approach for extracting coastal line feature in remote sensing. In: ISPRS Commission I Symposium on Integrated Remote Sensing at the Global, Regional and Local Scale.

[11] Matheron, G., 1975. Random Sets and Integral Geometry. Wiley, New York.

[12] Naegel, B., Passat, N., Ronse, C., February 2007. Grey-level hit-or-miss transforms - part 1: Unified theory. Pattern Recognition 40 (2), 635–647.

[13] Perret, B., Lefèvre, S., Collet, C., 2009. A robust hit-or-miss transform for template matching in very noisy astronomical images. Pattern Recognition 42 (11), 2470–2480.

[14] Puissant, A., Lefèvre, S., Weber, J., 2008. Coastline extraction in vhr imagery using mathematical morphology with spatial and spectral knowledge. In: XXIth ISPRS Congress - Benjing.

[15] Ronse, C., 1996. A lattice-theoretical morphological view on template extraction in images. Journal of Visual Communication and Image Representation 7 (3), 273–295.

[16] Serra, J., 1982. Image Analysis and Mathematical Morphology Vol I. Academic Press, London.

[17] Serra, J., 1988. Image Analysis and Mathematical Morphology Vol II: Theoretical advances. Academic Press, London.

[18] Soille, P., March 2002. Advances in the analysis of topographic features on discrete images. In: Proceedings of the 10th International Conference on Discrete Geometry for Computer Imagery DGCI'02. Vol. 2301 of Lecture Notes in Computer Sciences. pp. 175–186.

[19] Soille, P., 2003. Morphological Image Analysis : Principles and Applications. Springer, Berlin.

[20] Velasco-Forero, S., Angulo, J., 2010. Hit-or-miss transform in multivariate images. In: Advanced Concepts for Intelligent Vision Systems. Vol. 6474 of Lecture Notes in Computer Sciences. Springer Verlag, pp. 452–463.

[21] Weber, J., Lefèvre, S., July 2008. A multivariate hit-or-miss transform for conjoint spatial and spectral template matching. In: Image and Signal Processing, proceedings of the International Conference on Image and Signal Processing. Vol. 5099 of Lecture Notes in Computer Sciences. Springer Berlin/Heidelberg, pp. 226–235.

[22] Weber, J., Lefèvre, S., Gançarski, P., 2011. Interactive Video Segmentation based on Quasi-Flat Zones. In: Proceedings of IEEE International Symposium on Image and Signal Processing and Analysis. IEEE, France, pp. 265–270.