Journal of Mathematical Imaging and Vision 7, 359–374 (1997)

Typology of Spatial Structures of Images having the Same Color Set

ERNEST CHIARELLO
Laboratoire d'Ecologie des Eaux Douces et des Grands Fleuves (ESA CNRS 5023), Université Claude Bernard, 69622 Villeurbanne Cedex
ernest@ligia.univ-lyon1.fr

JEAN-MARC CHASSERY
Laboratoire TIMC-IMAG, Institut Albert Bonniot, Faculté de Médecine Domaine de la Merci, 38706 La Tronche Cedex

Abstract. A previous study proposed a new model for generating random spatial patterns for modelling the dispersion of different colors in an image (Chiarello et al., 1996). These simulations represented spatial structures in the sense of landscape ecology and they had to be compared to a real image. Thus, the aim of the present study was to measure the discrepancy between a set of simulations of multicolored mosaics and an observed pattern in order to build a Monte Carlo test. The multicolored mosaics were considered as random closed sets and described with the hitting function for all pairs of colors. This description provided large three-dimensional data tables (distances × color pairs × images) that were analyzed with the help of multiway data analyses. The partial triadic analysis was used. It provided a synthesis of the hitting function since the intrastructure enabled a typology of the distances for each image: the factorial coordinates of supplementary columns were plotted as ordinates against distances as abscissa. This synthetic descriptor provided a graphic tool for measuring the differences between the spatial dispersions of a same set of colors in several images.

Keywords: multicolored mosaics, spatial structures, Monte Carlo test, hitting function, multiway data analysis

1. Introduction

This paper deals with spatial statistics for image analysis. The proposed approach is to broaden the classical analyses (for which data are identically and independently distributed) to those that recognize the presence and importance of spatial information [8]. Such analyses are very important because satellites are amassing huge amounts of data, and only a small fraction of which is being analyzed [14]. Nevertheless, the class of patterns of two or more components forming a mosaic was until recently the least explored [26], so it is useful to propose new methods. The view is put on spatial statistics and not on pattern recognition and vision. It is based on the fact that data that are close together in space are often more alike than those that are far apart. The problems are principally ones of scale dealing up to more than a million data points (1024×1024) in each satellite image. The beginning of an "exploratory data analysis" for images has been already proposed [27], and we followed this way of investigation in this study.

Multicolored mosaics were considered as particular realizations of random closed sets [23]. The present paper proposed a method, which measured the dispersion of the components of the mosaics, in order to measure the discrepancy between simulations of mosaics (synthetic images) and observed spatial patterns (ecological maps in our case). The objective consisted in building a Monte Carlo test [10] for the multicolored mosaics. The method was based on well established tools, mathematical morphology and data analysis, but they were articulated to propose a global descriptor of the mosaics. Because the spatial generator of our multicolored mosaics has been described previously [7], the problem

of the mosaic generator is not discussed. We just recall that it is based on the stochastic pyramid [19, 24]. It should be noted that comparisons of images pixel by pixel and measures of correlation were rejected because they do not describe spatial structures of multicolored mosaics. The analysis was performed as follows. Firstly multicolored patterns were described with the hitting function used in the random set theory [23] for each pair of colors. In this way a Monte Carlo test for each pair of colors was built and a three-dimensional data set was obtained. Secondly a multiway data analysis, the partial triadic analysis [11, 30], was performed on the whole data set. It provided a synthetic descriptor of all the pairs of colors. A complete Monte Carlo test was then available for multicolored mosaics.

2. Describing Random Sets

From the theory of random sets [23, 28], the hitting function, which is a Choquet capacity, yields to functions that can measure the spatial dispersion of a component and the mutual associations between components in a given set, as for example the probability distribution function defines wholly a random variable. The theory of random sets is based on the hit-or-miss topology. It reflects the way image data are analyzed. It is well known that some sort of systematic probing is needed, which leads to the use of structuring elements B to check whether "B hits A" ($B \cap A \neq \emptyset$) or "B misses A" ($B \cap A = \emptyset$) [23]. A summary of the main definitions and results of random-set theory [20, 23] and mathematical morphology [28] is needed for introducing the method proposed here.

2.1. Hit-or-Miss Topology

Let $E = \mathbb{R}^2$ be a locally compact, Hausdorff, and separable space, and define $\mathcal{F}$ to be the set of all closed subsets of E, including the empty set $\emptyset$. Let $\mathcal{K}$ denote the set of all compact sets. For any $A \subset E$, define

$$\mathcal{F}^A = \{F \in \mathcal{F} \mid F \cap A = \emptyset\}$$

and

$$\mathcal{F}_A = \{F \in \mathcal{F} \mid F \cap A \neq \emptyset\}.$$

Then for any $K \in \mathcal{K}$ compact and any $G_1, \ldots, G_n$ open, the sets of form $\mathcal{F}^K \cap \mathcal{F}_{G_1} \cap \cdots \cap \mathcal{F}_{G_n}$ of $\mathcal{F}$ is a base for a topology on $\mathcal{F}$, called the hit-or-miss topology, and the topological space $\mathcal{F}$ is compact, Hausdorff, and separable. By taking countable unions and intersections of the open sets of the topological space $\mathcal{F}$, a σ-algebra $\sigma_{\mathcal{F}}$ on $\mathcal{F}$ is generated.

A random closed set is defined as a measurable mapping A from a probability space $(\Omega, \mathcal{A}, \mu)$ into the measurable space $(\mathcal{F}, \sigma_{\mathcal{F}})$. Let P be the law of A, induced on $\sigma_{\mathcal{F}}$ by $\forall \mathcal{V} \in \sigma_{\mathcal{F}},\ P(\mathcal{V}) = \mu(A^{-1}(\mathcal{V}))$.

In order to study any random set with the σ-algebra generated by $\mathcal{P}_G = \{X \in \mathcal{P}(\mathbb{R}^2) : X \cap G \neq \emptyset\}$, G open and $\mathcal{P}(\mathbb{R}^2)$ the set of all subsets of $\mathbb{R}^2$, it is equivalent to study its closure using the σ-algebra $\sigma_{\mathcal{F}}$. This is due to the equivalence $X \cap G \neq \emptyset$ if and only if $\bar{X} \cap G \neq \emptyset$.

It is important to find what important events make two random sets different. It is shown that it is not necessary to explore all compact K and all open $G_1, \ldots, G_n$. Indeed, for any $K \in \mathcal{K}$, the hitting (or capacity) function T_X is defined as $T_X(K) = P(X \in \mathcal{F}_K) = P(X \cap K \neq \emptyset)$. Then T_X has the following properties [23]:

1) $T_X(\Delta) = 0$ and $0 \leq T_X \leq 1$
2) T_X is increasing
3) T_X satisfies the following recurrence relations: for any $n \geq 0$, let $S_n(B_0; B_1, \ldots, B_n)$ denote the probability that X misses B_0 but hits $B_1, \ldots, B_n$; then:

$$\begin{aligned} S_0(B_0) &= 1 - T_X(B_0) \geq 0 \\ S_1(B_0; B_1) &= T_X(B_0 \cup B_1) - T_X(B_0) \geq 0 \\ S_n(B_0; B_1, \ldots, B_n) &= S_{n-1}(B_0; B_1, \ldots, B_{n-1}) \\ &\quad - S_{n-1}(B_0 \cup B_n; B_1, \ldots, \\ &\quad B_{n-1}) \geq 0 \end{aligned}$$

T_X is a Choquet capacity of infinite order. Conversely if a given T on $\mathcal{K}$ is a Choquet capacity of infinite order, then there exists a necessarily unique P_T on $\sigma_{\mathcal{F}}$ such that, $\forall K \in \mathcal{K},\ P_T(\mathcal{F}_K) = T(K)$ [20, 22].

A random closed set is wholly characterized by its hitting function $T(K)$. The problem of finding the ways to reduce the hitting-function test sets down from the full complement $\mathcal{K}$ is not discussed here.

2.2. Hitting Function and Functional

One can consider $P(A \cap K \neq \emptyset)$ or $P(A \cap K = \emptyset)$, i.e., the hitting function $T(K) = P(A \cap K \neq \emptyset) = P(\mathcal{F}_K)$, $\forall K \in \mathcal{K}$, or its functional $Q(K) = 1 - T(K)$. For K located at the origin O, the hitting function of a random closed set is $T(K) = P(A \cap K \neq \emptyset) = P(O \in A \oplus \check{K})$. If K is translated at any point x, the hitting function

becomes $T(K_x) = P\{K_x \subset A\} = P\{x \in A \oplus \check{K}\}$ and the functional is $Q(K_x) = P\{K_x \subset A^c\} = P\{x \in A^c \ominus \check{K}\}$.

It is necessary to have many realizations of the set A, and then to estimate its hitting function in each point x with an experimental frequency. Stationarity is generally assumed and, if A is ergodic, $T(K)$ is computed with a single realization [17] with, respectively, $T(K) = Area(A \oplus \check{K})$ or $Q(K) = Area(A^C \ominus \check{K})$, where *Area* is the area fraction, i.e., the area per unit area.

Each compact K is associated to some particular information about the set A which is considered as a realization of the random set.

2.3. *Multicomponent Sets*

When the set A is made of several components $(A_1, \ldots, A_n)$, it is characterized by a family of structuring elements $(K_1, \ldots, K_n)$; and the hitting function becomes:

$$\begin{aligned} &T(K_1, \ldots, K_n) \\ &\quad = P((K_1 \cap A_1 \neq \emptyset) \cap \cdots \cap (K_n \cap A_n \neq \emptyset)) \\ &\quad = P((K_1 \uparrow A_1) \cap \cdots \cap (K_n \uparrow A_n)) \\ &\quad = P(O \in \{A_1 \oplus \check{K}_1 \cap \cdots \cap A_n \oplus \check{K}_n\}) \\ &Q(K_1, \ldots, K_n) \\ &\quad = P((K_1 \cap A_1 = \emptyset) \cap \cdots \cap (K_n \cap A_n = \emptyset)) \\ &\quad = P\left(\left(K_1 \subset A_1^C\right) \cap \cdots \cap \left(K_n \subset A_n^C\right)\right) \\ &\quad = P\left(O \in \left\{A_1^C \ominus \check{K}_1 \cap \cdots \cap A_n^C \ominus \check{K}_n\right\}\right) \end{aligned}$$

Practically, mutual associations are studied between all pairs of components A_i and A_j. Two classes of points of view are possible, according to the fact that these morphological transformations are applied on a single or on two components [18, 29]. They can also be both presented using the symmetrical case, one of the structuring element being reduced to a single point.

Consequently, $T(K_i, K_j) = P(O \in \{(A_i \oplus \check{K}_i) \cap (A_j \oplus \check{K}_j)\})$ can be used, but the functional is not tractable since $P(O \in \{(A_i \ominus \check{K}_i) \cap (A_j \ominus \check{K}_j)\}) = 0$. A more sophisticated morphological transformation, the closing, is also used. Then the functional is $T(K_i, K_j) = P(O \in \{((A_i \oplus \check{K}_i) \ominus K_i) \cap ((A_j \oplus \check{K}_j) \ominus K_j)\})$.

Several components can be grouped together. The hitting function is analytically known [18] but it is better to compute a new image of the set A, the union of the components of interest being labelled with the same value.

The measurements depend on the structuring elements K_i and K_j. The bipoint $\{x, x + d\}$ and the closed ball $B_x(d)$ of radius d centred at x are the most commonly used. When the compacts are $(K_i = \{x\}, K_j = \{x\})$, the hitting function is expressed as $A_{ij} = Area(A_i \cap A_j)$, $\forall (i, j) \in [1, n]^2$. When $K_i = \{x\}$ and $K_j = \{x, x+d\}$, the crossed covariance between the two components is obtained: $C_{ij}(d) = Area(A_i \cap (A_j)_{-d})$, $\forall (i, j) \in [1, n]^2$. It should be noticed that the functional $Q(x, x+h)$ is the covariance $C(A^c, h)$ of A^c. We use the covariance of A because it is equivalent to study A or A^c.

Because the covariance has little discriminatory power [9, 27], the compacts $K_i = B_x(d)$ and $K_j = \{x\}$ are more usually used. Thus the morphological transformations propose a measure not only of the mutual association between components, but also of the shape of the considered components. The four following functions are measured:

$$\begin{aligned} \phi_{ij}(d) &= Area((A_i \oplus \check{B}(d)) \cap A_j) \quad \forall (i, j) \in [1, n]^2 \\ \theta_{ij}(d) &= Area(((A_i \oplus \check{B}(d)) \ominus B(d)) \cap A_j) \\ &\qquad \forall (i, j) \in [1, n]^2 \\ \psi_{ij}(d) &= Area((A_i \oplus \check{B}(d)) \cap (A_j \oplus \check{B}(d))) \\ &\qquad \forall (i, j) \in [1, n]^2 \\ \varphi_{ij}(d) &= Area(((A_i \oplus \check{B}(d)) \ominus B(d)) \cap (A_j \oplus \check{B}(d)) \\ &\qquad \ominus B(d))) \qquad \forall (i, j) \in [1, n]^2 \end{aligned}$$

These measures are particular realizations of the hitting function. They describe mutual associations between pairs of components. But they do not *characterize* a mosaic, because two different mosaics may have the same measurements. However, we hope that in practice these measures can discriminate real mosaic images. Consequently, each simulation of a multicolored mosaic can be described with hitting functions. We proposed to introduce these functions in a Monte Carlo test.

3. Monte Carlo Test with the Hitting Function

3.1. *Monte Carlo Test*

A Monte Carlo test is used when distribution theory is not manageable. It is based on the rank of an observed value of a statistic among a set of simulations. Let u_1

be the observed value of a statistic U, and let $u_{k/k=2,\ldots,t}$ be the corresponding values generated by independent random sampling from the distribution of U under a simple hypothesis $\mathcal{H}$ (complete spatial randomness, for example). Let $u_{(j)}$ denote the jth largest value amongst $u_{k/k=1,\ldots,t}$. Then under $\mathcal{H}$, $P\{u_1 = u_{(j)}\} = \frac{1}{t}$, $j = 1, \ldots, t$, and rejection of h on the basis that u_1 ranks hth largest or higher gives an exact, one sided test of size h/t [10].

Such a test is often used in spatial statistics when studying spatial mapped point patterns. Let $H(d)$ be a statistic of a spatial pattern. In the case of complete spatial randomness, the theoretical distribution $H(d)$ is assumed to be known. This statistic is estimated for each simulation $k, k = 2, \ldots, t$, and also for the observation $k = 1$ and is marked, respectively, $\hat{H}_{k/k=2,\ldots,t}(d)$ and $\hat{H}_1(d)$. Then, upper and lower simulation envelopes, $U(d) = \mathrm{Max}_{k=2,\ldots,t}\{\hat{H}_k(d)\}$ and $L(d) = \mathrm{Min}_{k=2,\ldots,t}\{\hat{H}_k(d)\}$ are defined. These envelopes can also be plotted against $H(d)$, and have the property that under complete spatial randomness, and for each d, $P\{\hat{H}_1(d) > U(d)\} = P\{\hat{H}_1(d) < L(d)\} = \frac{1}{t}$. They are intended to help in the interpretation of the plot of $\hat{H}(d)$ against d. Then it is possible to see for which distance d there is or not a discrepancy between simulations and observation.

Two possible approaches for the construction of an exact Monte Carlo test can be considered: firstly choose d_0 that corresponds to a particular and pertinent case for the problem, and then define $u_k = \hat{H}_k(d_0), \forall k, k = 1, \ldots, t$; or secondly define u_k to be a measure of the discrepancy between $\hat{H}_k(d)$ and $H(d)$, over the whole range of d as $u_k = \int\{\hat{H}_k(d) - H(d)\}^2 \mathrm{d}d, \forall k, k = 1, \ldots, t$. In each case, the test is based on the rank of u_1. The second has the advantage of objectivity, even if the comparison between curves is not trivial. When the theoretical distribution $H(d)$ is not known, a test can still be carried out if $H(d)$ is replaced by the estimation $\bar{H}_k(d) = \frac{1}{t-1}\sum_{s\neq k}\hat{H}_s(d)\ \forall k, k = 1, \ldots, t$. The u_k are no longer independent under complete spatial randomness, but the required property that all rankings of u_1 are equiprobable still holds [10]. So, the second approach has been chosen and applied to the hitting function.

3.2. A Data Analysis Problem

The hitting function yielded a set of descriptors for the random closed set A as functions of the distance d. Several types of analyses were available, depending on the choice of the structuring elements. Consequently, a Monte Carlo test could be performed for each couple of components (A_i, A_j), as described previously, using any of the statistics $C_{ij}(d), \phi_{ij}(d), \theta_{ij}(d), \psi_{ij}(d)$ and $\varphi_{ij}(d), \forall(i, j) \in [1, n]^2$. Let $H_{ij}(d)$ be one of them, computed for each simulation $k/k = 2, \ldots, t$, and also for the observation $k = 1$. The results of the Monte Carlo test were organized in a single table, in which each estimation $k = 1, \ldots, t$ was associated with an arbitrary color. Data images are shown in Fig. 1: three simulations (red, green, blue) and an observation (black). Figure 2 shows this table for the covariance function.

$$Tab\left(H_{ij}^k(d)\right) =$$

$$\begin{array}{c} \\ i \rightarrow \end{array} \begin{array}{c} \begin{array}{ccc} & j & \\ & \downarrow & \end{array} \\ \begin{pmatrix} H_{11}^k(d)_{k=1,\ldots,t} & \cdots & H_{n1}^k(d)_{k=1,\ldots,t} \\ \vdots & H_{ij}^k(d)_{k=1,\ldots,t} & \vdots \\ H_{1n}^k(d)_{k=1,\ldots,t} & \cdots & H_{nn}^k(d)_{k=1,\ldots,t} \end{pmatrix} \end{array}$$

Another table could be also calculated, which presented the observation, the mean of the simulation and the upper and lower envelopes of the Monte Carlo test for each pair of components (i, j): $Tab(H_{ij}^t(d), \bar{H}_{ij}^k(d), U(d), L(d))$. The scale of representation was specific for each (i, j) with a same scale, only the quantitatively more numerous components would be readable.

The discrepancies between simulations and reality for each (i, j) were graphically readable. But at the same time these results were too dispersed and did not provide a global response to the question: was the model good enough? So a synthesis of this information with the tools of data analysis was useful.

4. Synthesis of Data

The table $Tab(H_{ij}^t(d), \bar{H}_{ij}^k(d), U(d), L(d))$ could be considered as a three-dimensional data tables that could be studied with multiway data analyses. Many solutions for the analysis of such tables were available [21, 31]. But one of the simplest was the partial triadic analysis [12, 16, 30]. It is performed in three steps called interstructure, compromise and intrastructure (Fig. 3), which were recalled here. Firstly, it is necessary to present briefly the principal component analysis (PCA), which is fundamental for the partial triadic analysis. Then the partial triadic analysis is described, and the results are discussed.

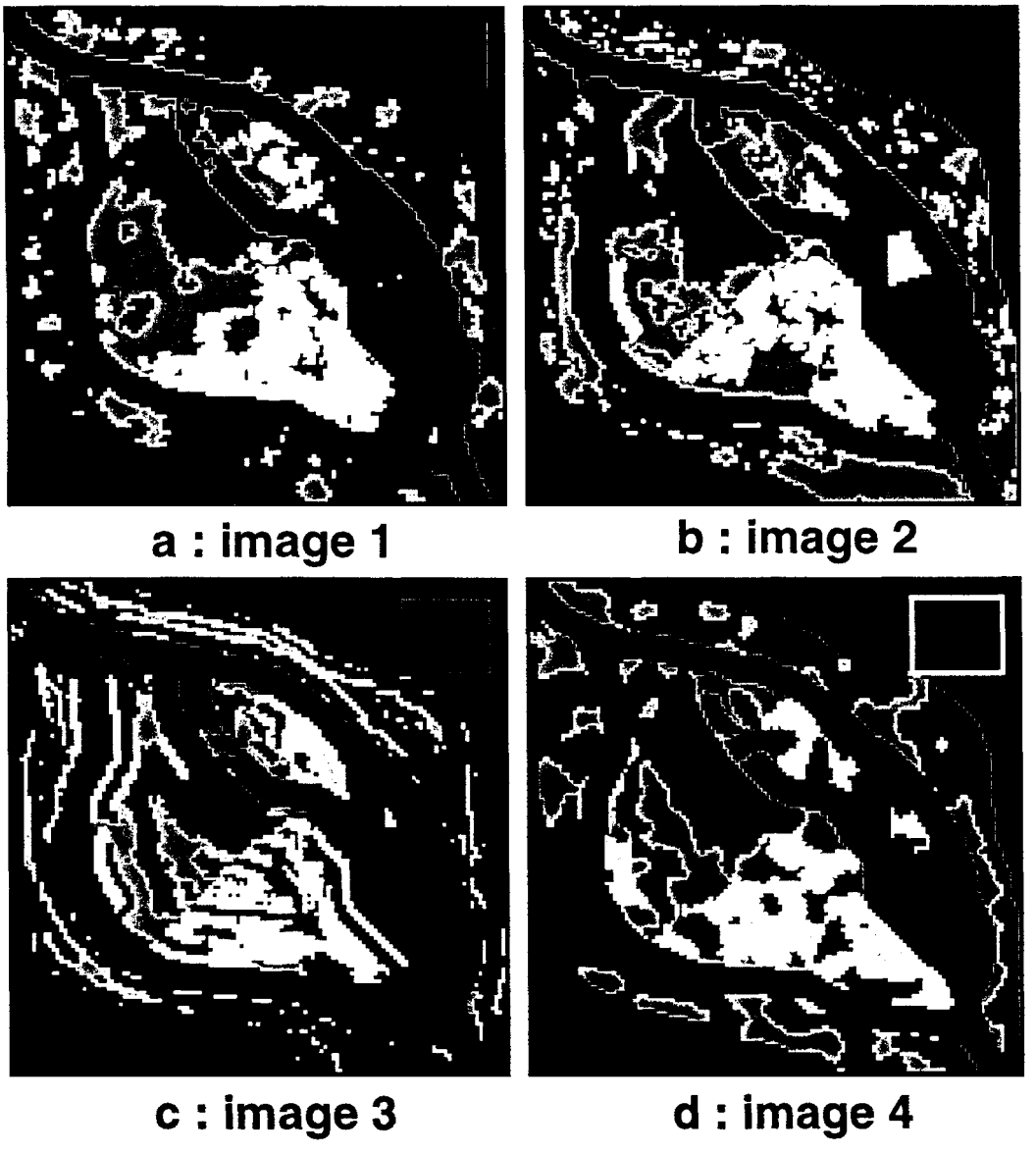

Figure 1. Four images (128×128) with different spatial structures. The first three images were the simulations, they were associated to colors (red, green and blue); and the fourth played the role of the observation (black). These four images had the same set of ten colors and had been simulated with a spatial generator of random patterns [7].

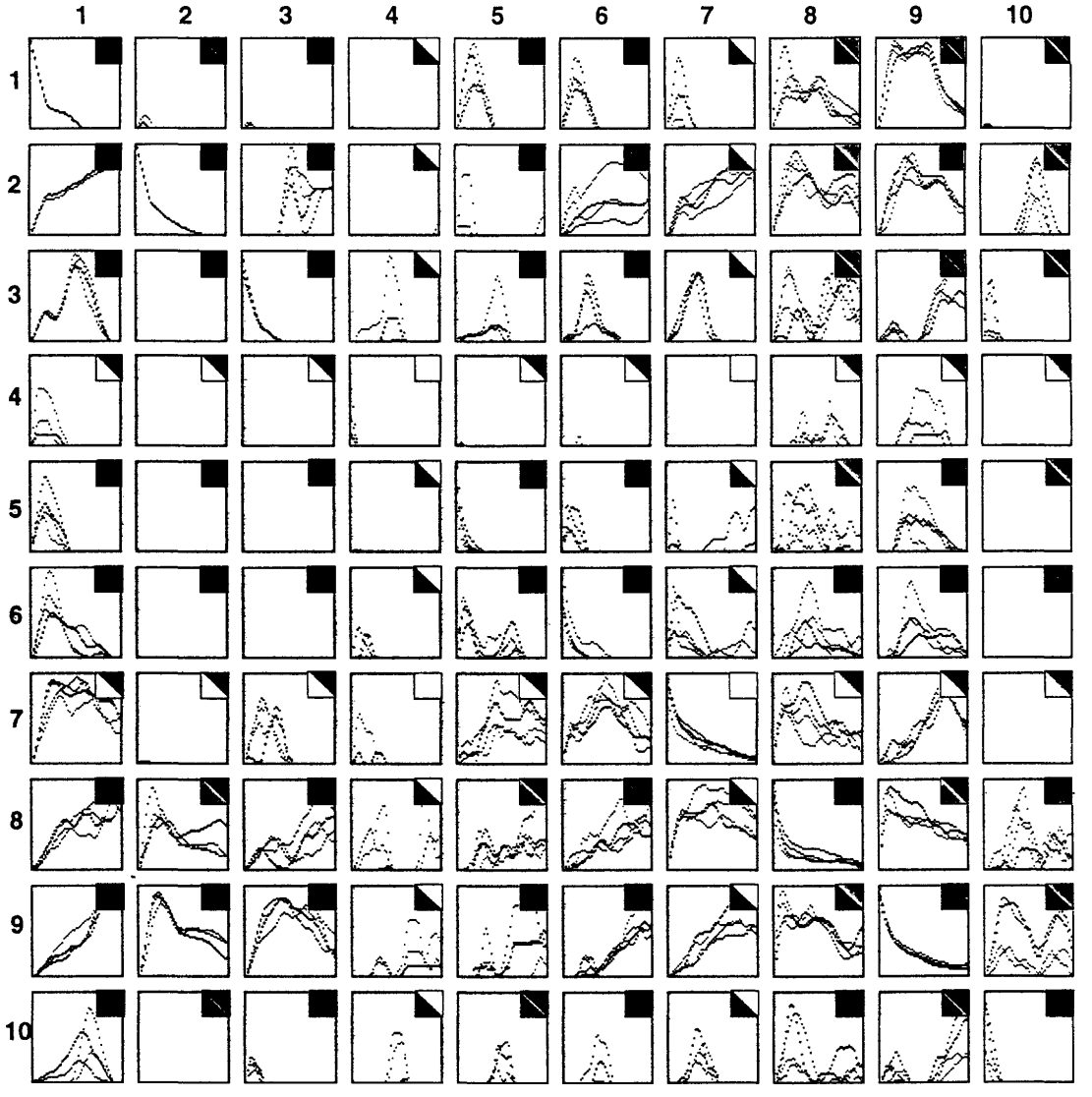

Figure 2. Table $Tab(H_{ij}^{k}(d))$ of the horizontal covariances $C_{ij}(d) = Area(A_i \cap (A_j)_{-d})$, $\forall (i, j) \in [1, n]^2$, $d \in [0, 50]$, $\alpha_d = 0^\circ$, $t = 4$, $n = 10$, computed on the 4 images of the Fig. 1.

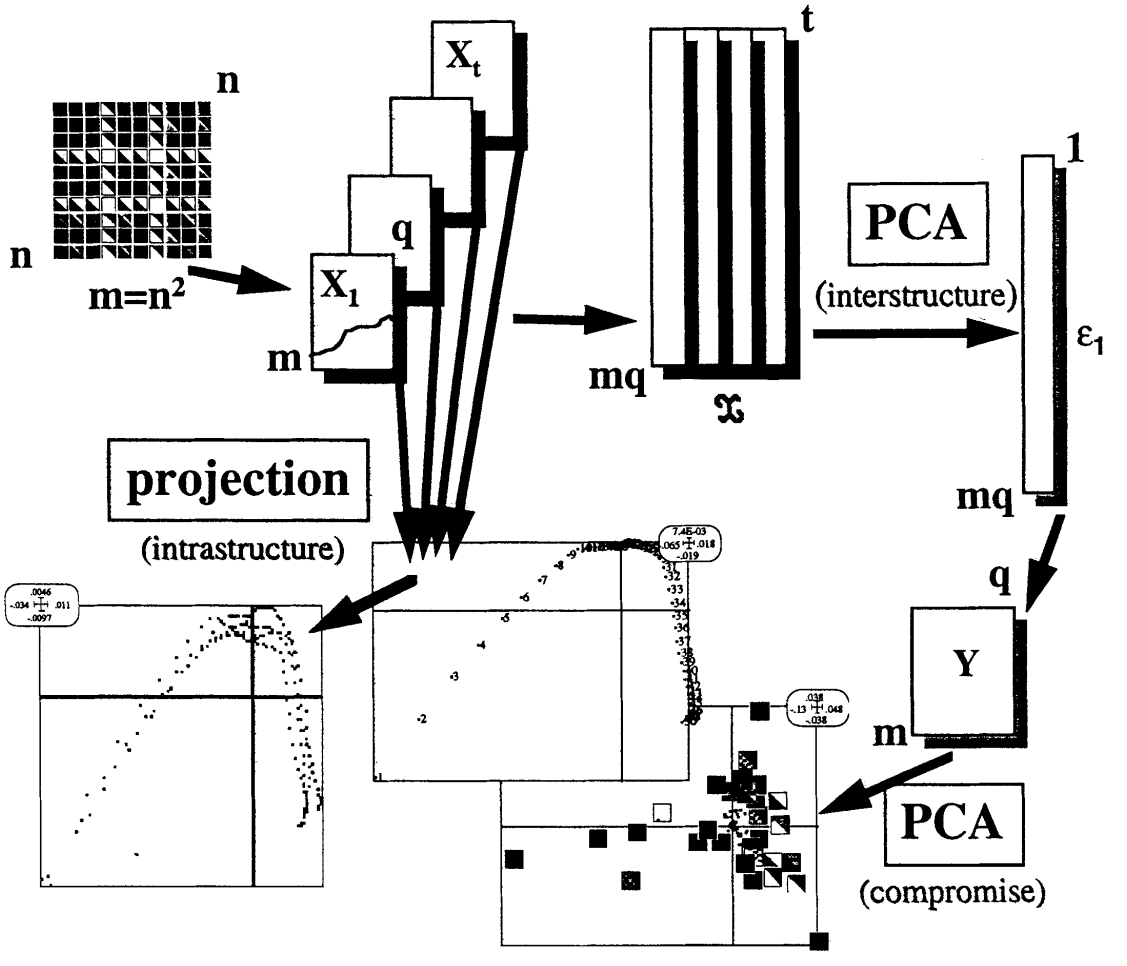

Figure 3. Schematic representation of the triadic partial analysis which was performed in three steps: interstructure, comprise and intrastructure. Initial data were tables of the horizontal covariances of Fig. 2. Only a PCA program was needed.

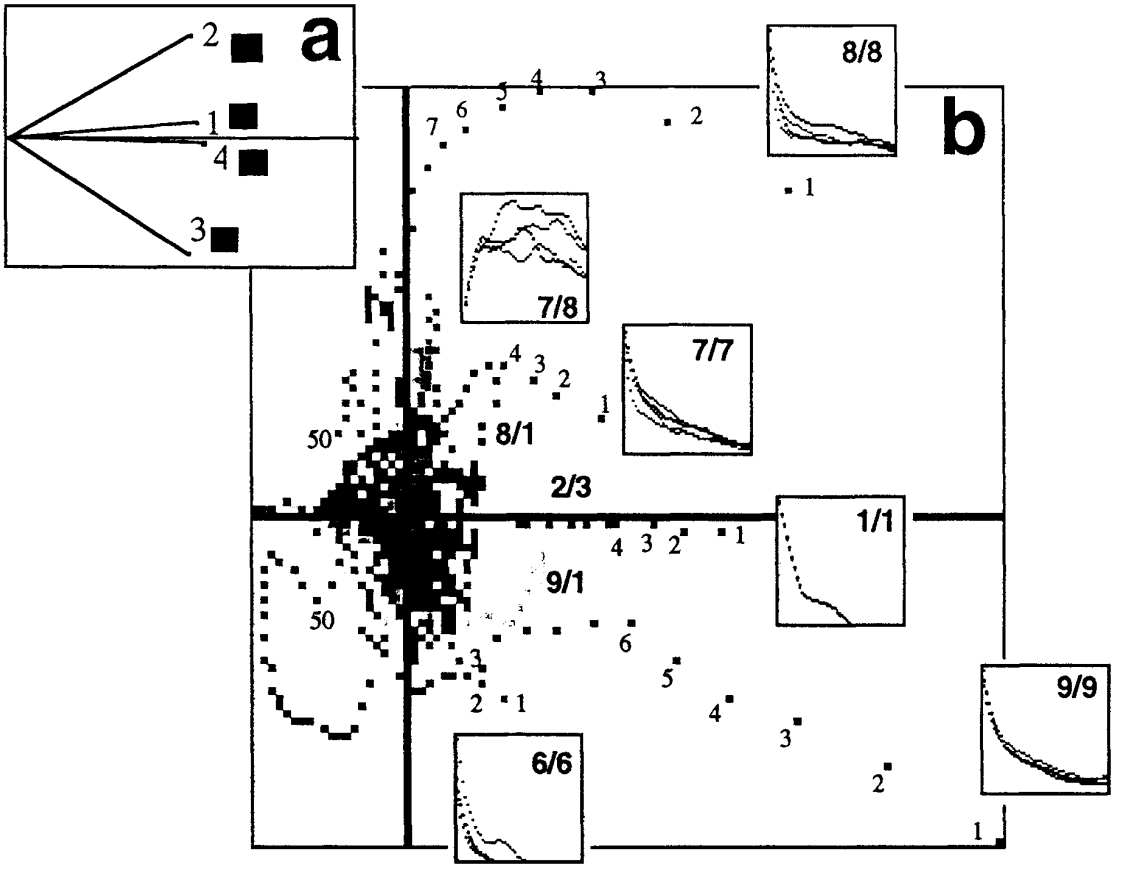

Figure 4. Factor maps generated by the interstructure: a) factor map of the columns of the interstructure showing a global description of the 4 images; b) factor map of the rows of the interstructure showing a description of the distances (1 to 50) relative to the typology of the images for each pair of colors (1/1 to 10/10); in addition some associated covariances were presented as small diagrams (numbered 1/1, 6/6, 7/7, 8/8, 7/8 and 9/9).

4.1. *Data Preparation*

We considered the table S_k ($n^2 = m$ pairs of components $\times$ q distances, $d \in [1, q]$) of size $(m \times q)$ for the simulation k:

$$S_k = \begin{pmatrix} H_{11}^k(1) & \cdots & H_{11}^k(q) \\ \vdots & & \vdots \\ H_{nn}^k(1) & \cdots & H_{nn}^k(q) \end{pmatrix} \quad \forall k \in [1, t]$$

In order to normalize the tables, the data were first computed into percentages per column, providing the table Z_k, $\forall k \in [1, \ldots, t]$:

$$Z_{ij}^k(d) = \frac{H_{ij}^k(d) \times 100}{\sum_{i=1,j=1}^{n,n} H_{ij}^k(d)} \quad \forall d \in [1, q],$$

Each column of Z_k was therefore a probability distribution of the pairs of colors. Then each Z_k was centred by rows to move each gravity center of each Z_k to the origin O; the data matrices $X_{k/k=1,\ldots,t}$ obtained were:

$$X_{ij}^k(d) = Z_{ij}^k(d) - \frac{1}{t \cdot q} \sum_{k=1}^{t} \sum_{d=1}^{q} Z_{ij}^k(d)$$
$$\forall (i, j) \in [1, n]^2$$

4.2. *Principal Component Analysis*

For the analysis of a data table X (m rows and q columns), the method uses a linear projection algorithm, which maps the set of m q-dimensional patterns onto a g-dimensional space, where $g < q \cdot g = 2$ is usually chosen because this permits visual examination of the initial multivariate data. Two diagonal matrices of weights are defined for rows and columns, $Q_m = \frac{1}{m} I_m$ and $Q_q = \frac{1}{q} I_q$, in the canonical basis (e_i) and (f_j); these matrices correspond to metrics in the two Euclidean spaces $\mathbb{R}^m$ and $\mathbb{R}^q$; I_m and I_q are the associated unity matrices, the weights here are uniform. The scalar product in the sense of Q_m between two centred variables x_0 and y_0 equals the covariance between the variable x and y: $(x_0 \mid y_0)_{Q_m} = \mathrm{cov}(x, y)$.

The principal component analysis of the triplet (X, Q_m, Q_q) leads to the principal axes $\varepsilon = [\varepsilon_1, \ldots, \varepsilon_q]$ in $\mathbb{R}^q$, which are the Q_q-normed eigenvectors of $X' Q_m X Q_q$. The new coordinates of the rows in ε are the principal components: $\varphi = [\varphi_1, \ldots, \varphi_q]$: $N = X Q_q \varepsilon$. In the same way, the columns have new coordinates: $P = X' Q_m \varphi$ where $\varphi_k = \frac{1}{\sqrt{\lambda_k}} X Q_q(\varepsilon_k)$ $\forall k, k = 1, \ldots, m$.

The separate analysis of each table X_k was out of interest for our question because we were looking for an analysis that took into account all the tables $X_{k/k=1,\ldots,t}$ simultaneously. The central theorem of the partial triadic method allowed that, but a new basis for the tables had to be defined.

4.3. *New Basis*

Each X_k belonged to $\mathcal{L}(\mathbb{R}^{t^*}, \mathbb{R}^m)$, called the space of tables, and where $\mathbb{R}^{t^*}$ was the algebraic dual space of $\mathbb{R}^t$. Thus there existed a linear application f between $\mathbb{R}^{t^*}$ and $\mathcal{L}(\mathbb{R}^{t^*}, \mathbb{R}^m)$, (e_i^*) being the basis of $\mathbb{R}^{t^*}$, defined as [12, 16]:

$$f : \mathbb{R}^{t^*} \to \mathcal{L}(\mathbb{R}^{t^*}, \mathbb{R}^m)$$
$$e_s^* \mapsto f(e_s^*) = X_k$$

The space of tables $\mathcal{L}(\mathbb{R}^{t^*}, \mathbb{R}^m)$ was canonically isomorphic to $\mathbb{R}^{t^*} \otimes \mathbb{R}^m$, and it could be supplied with the basis of the vectors $e_i^* \otimes f_j$; consequently each X_k could be written in a single column [12]. Therefore, the data matrix $X_{k/k=1,\ldots,t}$ of size $(m \times q \times t)$ could be transformed in a single matrix $\mathcal{X}$ with $m \times q$ rows and t columns. This matrix was associated to the linear application f.

The new data matrix $\mathcal{X}$ was:

$$[\mathcal{X}_{ij}] = \sum_{i=1}^{m} \sum_{j=1}^{q} \mathcal{X}_{ij} f_{i \otimes j} = \begin{pmatrix} X_{111} & & X_{t11} \\ \vdots & \cdots & \vdots \\ X_{11q} & & X_{t1q} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ X_{1m1} & & X_{tm1} \\ \vdots & \cdots & \vdots \\ X_{1mq} & & X_{tmq} \end{pmatrix}$$

The scalar product in $\mathbb{R}^t$ was $(\cdot \mid \cdot)_\Omega$; its matrix was $\Omega = I_t$. In $\mathcal{L}(\mathbb{R}^{t^*}, \mathbb{R}^m)$, the scalar product was $(\cdot \mid \cdot)_\Delta$ and was defined by its matrix $\Delta = \frac{1}{m} I_m \otimes \frac{1}{q} I_q = \frac{1}{mq} I_{mq}$ in the basis $(f_{i \otimes j})$.

The first step of a triadic partial method was thus reduced to the non centred principal component analysis of the triple $(\mathcal{X}, I_t, \frac{1}{mq} I_{mq})$: it was called interstructure (Fig. 3). An overall description of the images was obtained.

4.4. Compromise and Intrastructure

Under the constraint $\|\varepsilon\|_{I_t}^2 = 1$, the interstructure (Fig. 3) provides a first eigenvector ε_1 that maximizes $\|\mathcal{X}Q(\varepsilon)\|_{\Delta}^2 = (\mathcal{X}Q(\varepsilon) \mid \mathcal{X}Q(\varepsilon))_{\Delta} = Trace(Y'DYQ) = \frac{1}{mq} Trace(Y'Y)$, which was the inertia of a new table, marked Y, of size $(m \times q)$. This means that the table $\mathcal{X}$, juxtaposition of the tables $X_{k/k=1,\ldots,t}$, was summarized in a single table Y which was by construction the most representative, because the new statistical triple $(Y, \frac{1}{m} I_m, \frac{1}{q} I_q)$ had the most important inertia. This is the reason why this step is called compromise. The table Y was a linear combination of the initial tables $X_{k/k=1,\ldots,t}$, where the coefficients were precisely the components of the first eigenvector $\varepsilon_1 = (a_1, \ldots, a_t) : Y = \sum_{k=1}^{t} a_k X_k$. Y was also analyzed with a PCA. It should be noticed that the interstructure needs the principal component analysis of the compromise as optimal. The analysis of the compromise enabled the simultaneous representation of distances and pairs of colors.

The last step of the analysis was the intrastructure (Fig. 3), which consisted as a double projection of the rows and columns of the initial tables $X_{k/k=1,\ldots,t}$ as supplementary individuals and supplementary variables. A typology of distances for each image and a typology of each pair of colors for each image were then computed.

4.5. Results

The interstructure presented a new space in which 95% ($\lambda_1 = 0{,}4567 \cdot 10^{-3}$, $\lambda_2 = 0{,}227 \cdot 10^{-3}$), of the variance was retained after PCA of the triple $(\mathcal{X}, I_t, \frac{1}{m_q} I_{mq})$. The data were strongly structured (autocorrelation of the distances). The projection of columns of the table $\mathcal{X}$ (images) gave a global description of them (Fig. 4(a)). Image 1 was only close to image 4 on the factor map, whereas images 2 and 3 were distant.

4.5.1. Description of Colors. The projection of the rows of $\mathcal{X}$ indicated for which pairs of colors the proximities and oppositions were important (Fig. 4(b)). A description of the distances in relation to the typology of images was thus obtained for each pair of colors.

The first factor corresponded to a classical size effect (Fig. 4(b)), which related the overall structure of the data. The most abundant colors in the images were successively 9 and 8, 1, and finally 7. Consequently, their covariances were high, particularly for the short distances. The factor F1 opposed the crossed covariances to the covariances: the first were increasing, see 7/8 for example, although the second were decreasing, 8/8 or 9/9. One should remark that the distances associated to 9/9 or 8/8 cross the factorial map, in the opposite sense of the factor F1. This related to the fact that for the higher distances, the covariances were similar to the crossed covariances since they had quite similar quantities of pixels.

The second factor indicated oppositions and proximities between images, and which pairs of colors were concerned. The image 2 differed from the image 3 because its covariance for the pair of colors 8/8 had higher values. In a similar way, the image 3 differed from the others because its covariances for the pairs of colors 9/9 and 6/6 were different from the other images.

The non centred PCA of $(Y, \frac{1}{m} I_m, \frac{1}{q} I_q)$ proposed, with an inertia of 95%, a synthetic summary ($\lambda_1 = 3{,}766 \cdot 10^{-4}$, $\lambda_2 = 0{,}5842 \cdot 10^{-4}$) of the pairs of colors with an interpretation key given by the map of the distances (Fig. 5). The projection of the columns of the

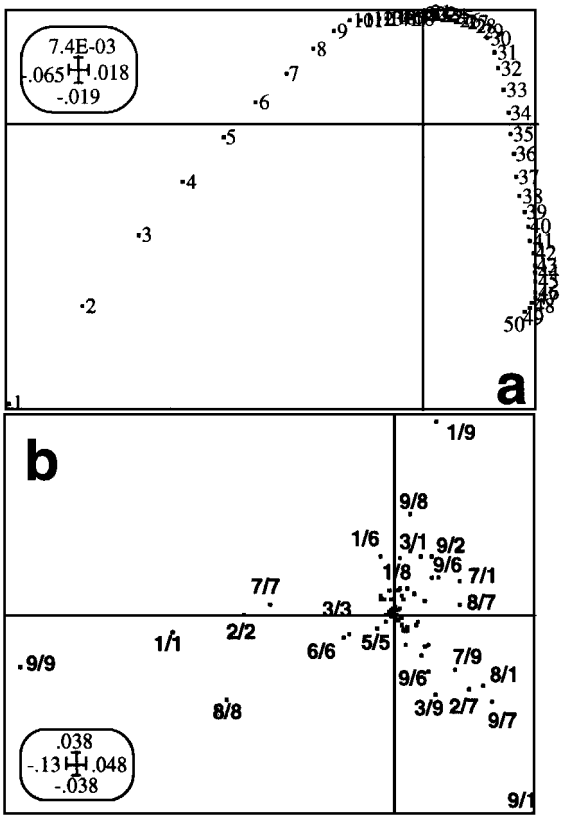

Figure 5. Factor maps generated by the compromise: a) factor map of the columns of the compromise corresponding to an averaged typology of the distances (1 to 50); b) factor map of the rows of the compromise corresponding to an averaged typology of the pairs of colors (1/1 to 10/10).

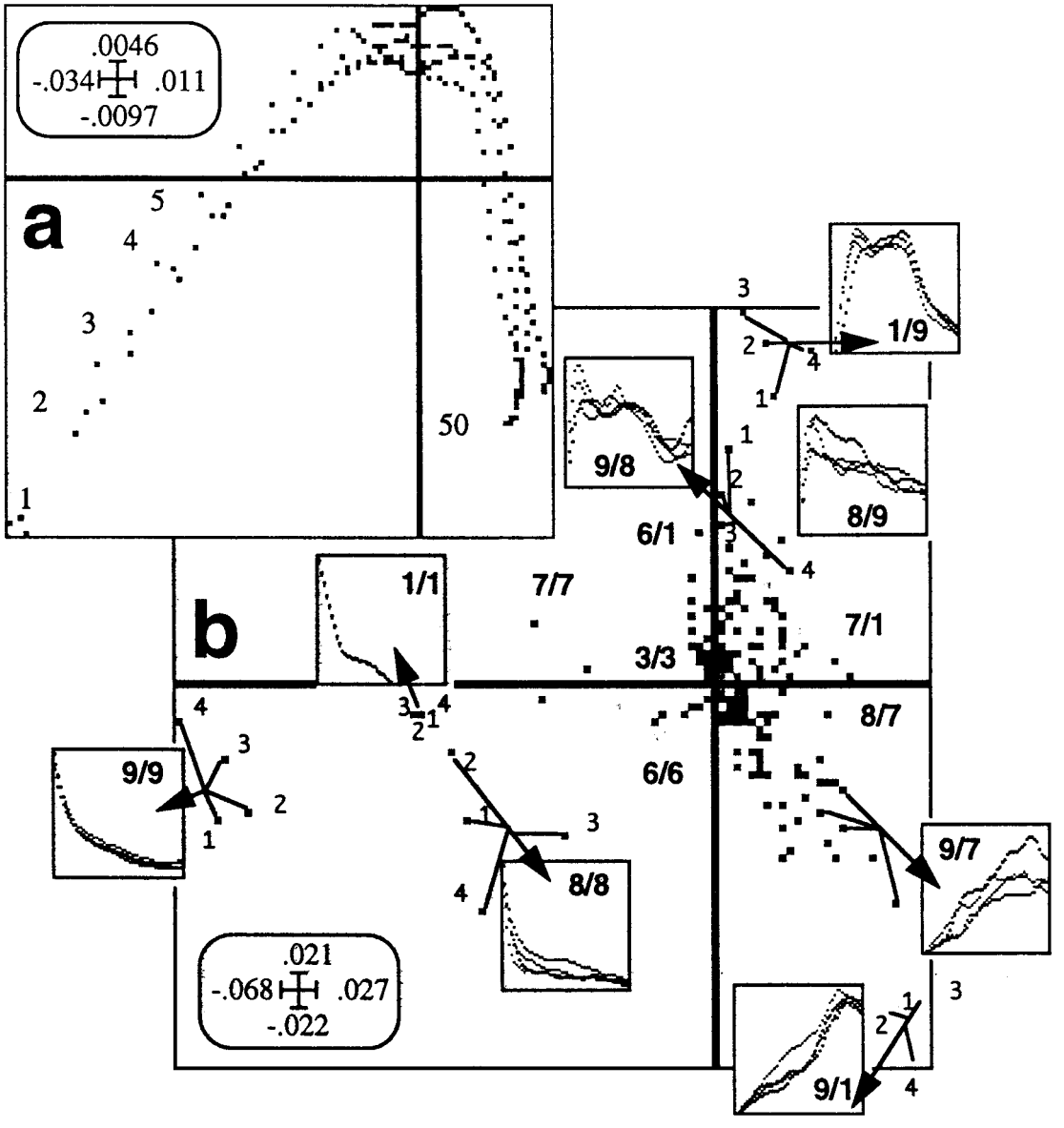

Figure 6. Factor maps generated by the intrastructure: a) typology of the distances (1 to 50) for the 4 images showing the discrepancy between images for all the distances; b) typology of the pairs of colors for the 4 images, with some associated covariances presented as small diagrams (1/1, 1/9, 8/8, 8/9, 9/1, 9/7, 9/8, 9/9).

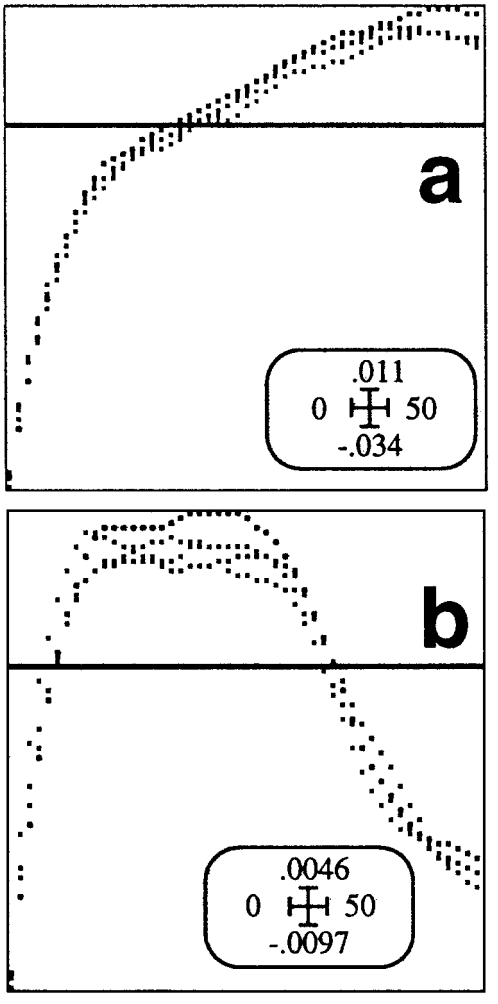

Figure 7. Synthetic covariance: a) with the axis F1 of the intrastructure; b) with the axis F2. Both were global descriptions of the spatial structures of the images for all the distances: b) related the discrepancy between the spatial structures of the 4 images while (a) indicated they had the same colors.

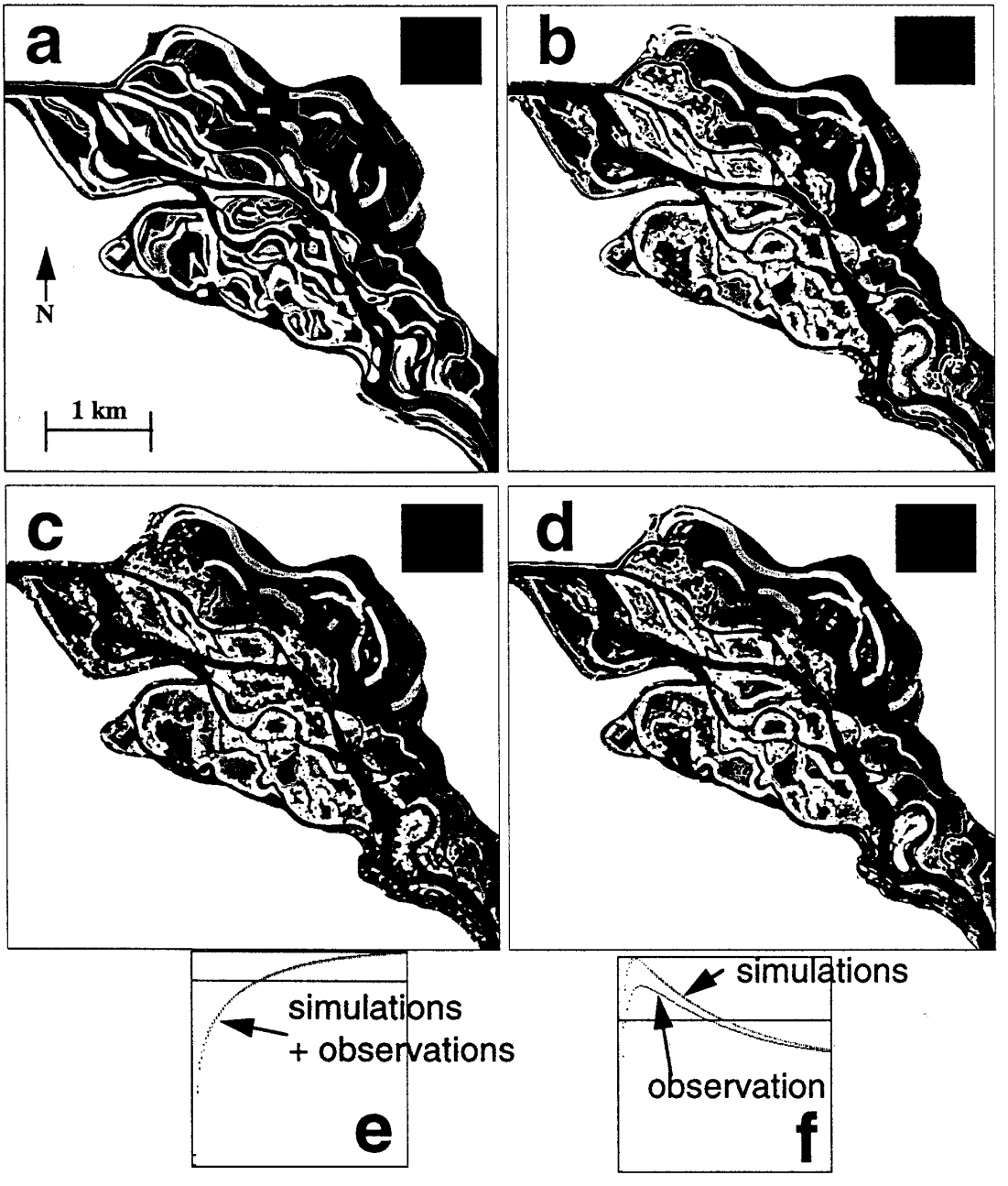

Figure 8. Application of the method on a larger image (850 × 900) of the Rhône River floodplain (a, black) at Brégnier-Cordon (France) (from Pautou et al. [25] and Bravard et al. [3]). Each color of the map represented an ecological unit; the simulations (b, c, d: red, green and blue) were computed with a controlled version of the stochastic pyramid from Meer and Connelly [24] and (Chiarello et al. [6]). The descriptor $H_{ij}(d)$ of the mosaics was the dilatation $\phi_{ij}(d) = Area((A_i \oplus \check{B}(d)) \cap A_j)$, $\forall(i, j) \in [1, n]^2$, $d \in [0, 100]$, $t = 4$, $n = 15$. The results of the Monte Carlo test are shown on diagrams of synthetic dilatation: (e) with the axis F1 and (f) with the axis F2. One can note (f) the discrepancy between simulations (red, green and blue curves) and reality (black curve).

compromise gave indeed an average typology of the distances, which completely splited the very short distances from the others, and which showed the strong autocorrelation between distances (Fig. 5(a)). It would be better to remove such a structuring effect from the data because it was trivial. In spite of this weakness, the typology of the distances allowed the interpretation of the averaged typology of the pairs of colors, obtained after projection of the rows of the compromise (Fig. 5(b)). The factors then had the same meanings as those of the interstructure. These maps enabled the interpretation of the intrastructure.

The intrastructure (Fig. 6) represented the more interesting step of the analysis in our case, because we obtained two synthetic descriptions of the initial tables: a typology of the distances and a typology of the pairs of colors for each image, the first being the interpretation key of the second, and *vice versa*.

The typology of the colors for each image was obtained by projection of the initial tables as supplementary rows in the PCA of the compromise (Fig. 6(a)). We could see the images for each pair of colors. The interpretation of the factors did not change. The factor F1 opposed pairs of similar colors (covariances) to pairs of different colors (crossed covariances). The factor F2 related proximities and differences between images. The typology of the distances was an interpretation key because it indicated for which distances the differences between images were important.

The typology of the distances for each image was obtained by projection of the initial tables as supplementary columns in the PCA of the compromise (Fig. 6(b)). In that case, we could see the images for each distance. The key given by the typology of the pairs of colors showed the ones for which the differences between images were the most important. Image 3 differed from the others for the short distances (from 1 to 10), because of its streamlined shapes. Image 4 was different for greater distances (from 10 to 30) because it was smoothed, and consequently, its regions were more homogeneous and rounder.

4.5.2. Synthetic Description. The last step consisted in the synthesis of the hitting function for a Monte Carlo test. Since the intrastructure enabled a typology of the distances for each image, it was possible to plot the factorial coordinates of the supplementary columns as ordinate against the distance as abscissa (Fig. 7). Two graphs were obtained, one on F1 (Fig. 7(a)) and the other on F2 (Fig. 7(b)). A synthetic descriptor was created which could be called "synthetic covariance" when $H_{ij}(d) = C_{ij}(d)$. Factor F1 indicated that the images had globally the same colors and quite similar structures. But the factor F2 gave a measure of the discrepancy between images: the synthetic covariance 4 differed from the others (smoothed image), while the synthetic covariance 3 was very particular (streamlined shapes) (Fig. 7(b)).

5. Conclusion

The proposed solution integrated in a genuine Monte Carlo test some basic tools of image analysis, random set theory and data analysis. A multicolored mosaic was then seen as a tessellation of random closed sets, and was consequently analysed with the hitting function. The data were synthetized with the triadic partial analysis. A visual measure of the distance between spatial structures of simulations and observation was finally obtained. The discrepancy between a real map of the vegetation, the Rhône River floodplain in Brégnier-Cordon (France) (Fig. 8(a)) and its associated models (Fig. 8(b), (c) and (d); [7] was then quantified (Fig. 8(e) and (f). It is clear that the model did not fit the data (Fig. 8(f)).

Such analyses for spatial data are very necessary. We recalled above that satellites are amassing a lot of data. These data are landscape images that have to be compared. It is important to measure the changes between successive dates, but also to give a sense to these changes. It can be shown indeed that image at time t does not match to image at time $t+1$. But the images have perhaps the same spatial structures, because landscapes are above all shifting mosaics. When such situations occur, new challenges are proposed to conceptors of geographic information systems [4, 15] because landscape have not to be recorded as fixed mosaics but as shifting mosaics [13]. Consequently dynamic spatial models [1, 2, 6, 8] are necessary.

Appendices

Translation:	$A_h = A \oplus h$
	$= \{a + h \mid a \in A\}$
Transposition:	$\check{A} = \{-a \mid a \in A\}$
Multiplication by a scalar:	$\lambda A = \{\lambda a \mid a \in A\}$
Complementation:	$A^C = \{x \in \mathbb{R}^2 : x \notin A\}$

Minkowski addition:

$$A \oplus B = \{x \in \mathbb{R}^2, \exists b \in B, x - b \in A\}$$
$$= \{a + b \mid a \in A, b \in B\}$$

Minkowski substraction:

$$A \ominus B = \{x \in \mathbb{R}^2, \forall b \in B, x - b \in A\}$$
$$= (A^C \oplus B)^C$$
$$= \cap \{A \oplus b : b \in B\}$$

Erosion: $E^B(A) = A \ominus \check{B}$
$$= \{x \in \mathbb{R}^2 : B_x \subseteq A\}$$

Dilatation: $D^B(A) = A \oplus \check{B}$
$$= \{x \in \mathbb{R}^2 : A \cap B_x \neq \emptyset\}$$

Opening: $O^B(A) = (A \ominus \check{B}) \oplus B$

Closing: $F^B(A) = (A \oplus \check{B}) \ominus B$

Acknowledgments

We gratefully acknowledge D. Chessel for valuable comments on a french version of the manuscript, M. Schmitt and two anonymous referees for greatly improving an earlier draft of the manuscript.

References

1. N. Ahuja and B.J. Schachter, *Pattern Models*, Wiley: New York, 1983, p. 309.
2. W.L. Baker, "A review of models of landscape change," *Landscape Ecology*, Vol. 2, pp. 111–133, 1989.
3. J.P. Bravard, C. Amoros, and G. Pautou, "Impact of civil engineering works on the successions of communities in a fluvial system," *Oikos*, Vol. 47, pp. 92–111, 1986.
4. P.A. Burrough, *Principles of Geographical Information Systems for Land Resource Assessment*, Clarendon Press: Oxford, 1986.
5. E. Chiarello, C. Amoros, G. Pautou, and J.M. Jolion, "Landscapes modelling using the stochastic pyramid and analytical models of successions," *Ecological Modelling* (submitted).
6. E. Chiarello, J.M. Jolion, and C. Amoros, "Model of spatial organization of vegetation in river floodplains," *Acta Stereologica*, Vol. 12, No. 2, pp. 255–261, 1993.
7. E. Chiarello, J.M. Jolion, and C. Amoros, "Regions growing with the stochastic pyramid: Application in landscape ecology," *Pattern Recognition*, Vol. 29, No. 1, pp. 61–75, 1996.
8. N.A.C. Cressie, *Statistics for Spatial Data*, Wiley & Sons: New York, 1991, p. 900.
9. P.J. Diggle, "Binary mosaics and the spatial patterns of heather," *Biometrics*, Vol. 37, pp. 531–539, 1981.
10. P.J. Diggle, *Statistical Analysis of Spatial Point Patterns*, Academic Press: London, 1983, p. 148.
11. S. Dolédec and D. Chessel, "Recent developments in linear ordination methods for environmental sciences," *Advances in Ecology*, Vol. 1, pp. 133–155, 1991.
12. F. Glaçon, "Analyse conjointe de plusieurs matrices de données. Comparaison de différentes méthodes," Thèse de 3° cycle, Scientific and Medicale University of Grenoble, p. 73, 1981.
13. D.C. Glenn-Lewin and E. van der Maarel, "Patterns and processes of vegetation dynamics," in *Plant Succession: Theory and Prediction*, Glenn-Lewin et al. (Eds.), Chapman & Hall: London, 1992, pp. 11–59.
14. M.F. Goodchild, "Integrating GIS and remote sensing for vegetation analysis and modeling: methodological issues," *J. of Vegetation Sci.*, Vol. 5, No. 5, pp. 615–626, 1994.
15. M.F. Goodchild, B.O. Parks, and L.T. Steyaert, *Environmental Modeling with GIS*, Oxford University Press: New York, 1993.
16. P.A. Jaffrenou, "Sur l'analyse des familles finies de variables vectorielles: bases algébriques et application à la description statistique," Thèse de 3° cycle, C. Bernard Lyon I University, p. 97, 1978.
17. D. Jeulin, "Modèles morphologiques de structures aléatoires et de changements d'échelle," Thèse d'Etat, Caen University, p. 261, 1991.
18. D. Jeulin, "Study of spatial distributions in multicomponent structures by images analysis," *Acta Stereologica*, Vol. 5, No. 2, pp. 233–239, 1986.
19. J.M. Jolion and A. Rosenfeld, *A Pyramid Framework for Early Vision*, Kluwer Academic Publishers: London, 1994, p. 218.
20. D.G. Kendall, "Foundations of a theory of random sets," in *Stochactic Geometry*, E.F. Harding and D.G.Kendamm (Eds.), Wiley: New York, 1974, pp. 322–376.
21. P. Kroonenberg, *Three-Mode Principal Component Analysis*, DSWO Press, 1983, p. 398.
22. G. Matheron, "Random sets theory and its application to stereology," *Journal of Microscopy*, Vol. 95, pp. 15–23, 1971.
23. G. Matheron, *Random Sets and Integral Geometry*, Wiley: New York, 1975.
24. P. Meer and S. Connelly, "A fast parallel method for synthesis of random patterns," *Pattern Recognition*, Vol. 22, pp. 189–204, 1989.
25. G. Pautou, J. Girel, B. Lachet, and G. Aïn, "Recherches écologiques dans la vallée du Haut-Rhône français," *Documents de Cartographie Ecologique*, Vol. 22, pp. 5–63, 2 appended maps, 1979.
26. B.D. Ripley, *Spatial Statistics*, Wiley: New York, 1981, p. 252.
27. B.D. Ripley, *Statistical Inference for Spatial Processes*, Cambridge University Press: Cambridge, 1988, p. 148.
28. J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press: London, 1982, p. 610.
29. D. Stoyan and J.Ohser, "Correlation between planar random structures (with an ecological example)," *Biometrical J.*, Vol. 24, pp. 631–747, 1982.
30. J. Thioulouse and D. Chessel, "Les analyses multitableaux en écologie factorielle. I: De la typologie d'état à la typologie de fonctionnement par l'analyse triadique," *Acta Oecologica, Oecologica General*, Vol. 8, No. 4, pp. 463–480, 1987.
31. L.R. Tucker, "Implications of factor analysis of three-way matrices for measurement of change," in *Problems in Measuring Change*, C.W. Harris (Ed.), University of Wisconsin Press: Madison, 1963, pp. 122–137.

Ernest Chiarello received a Ph.D. in Biometrical Sciences from University Claude Bernard Lyon 1, France in 1994. His research focuses on landscape ecology and image analysis for the modelling of spatial structures and temporal dynamics of multicolored mosaics. He developed models that integrates ecological rules (successions, strategies) for prediction of distribution of ecological units in river floodplains at different scales. He is presently contractual teacher at the Institut National des Sciences Appliquées of Lyon, France.

Jean-Marc Chassery was born in France on April 1949. He received the Doctorat d'Etat degree in 1984. He is currently Director of Research at CNRS and manages the image group INFODIS at TIMC-IMAG Laboratory in Grenoble, France. His primary research interests are digital geometry and image processing including multiscale and adaptive spatial partitioning in 2D and 3D. Such models have been applied in image reconstruction, segmentation and compression.