INTERNATIONAL COMPUTER SCIENCE INSTITUTE

1947 Center St. • Suite 600 • Berkeley, California 94704-1198 • (510) 643-9153 • FAX (510) 643-7684

# Image segmentation through contextual clustering

A. Baraldi*, P. Blonda†, F. Parmiggiani‡ and G. Satalino§

TR-98-009

June 1998

**Abstract.** Several interesting strategies are adopted by the well-known Pappas clustering algorithm to segment smooth images. These include exploitation of contextual information to model both class conditional densities and *a priori* knowledge in a Bayesian framework. Deficiencies of this algorithm are that: i) it removes from the scene any genuine but small region; and ii) its feature-preserving capability largely depends on a user-defined smoothing parameter. This parameter is equivalent to a clique potential of a Markov Random Field model employed to capture known stochastic components of the labeled scene. In this paper a modified version of the Pappas segmentation algorithm is proposed to process smooth and noiseless images requiring enhanced pattern-preserving capability. In the proposed algorithm: iii) no spatial continuity in pixel labeling is enforced to capture known stochastic components of the labeled scene; iv) local intensity parameters, pixel labels, and global intensity parameters are estimated in sequence; and v) if no local intensity average is available to model one category in the neighborhood of a given pixel, then global statistics are employed to determine whether that category is the one closest to pixel data. Results show that our contextual algorithm can be employed: vi) in cascade to any noncontextual (pixel-wise) hard $c$-means clustering algorithm to enhance detection of small image features; and vii) as the initialization stage of any crisp and iterative segmentation algorithm requiring priors to be neglected on earlier iterations (such as the Iterative Conditional Modes algorithm).

**Keywords**: Markov Random Field, Bayes' theorem, image segmentation.

*ICSI, 1947 Center Street, Suite 600, Berkeley, CA 94704-1198, baraldi@icsi.berkeley.edu
†IESI-CNR, Via Amendola 166/5, 70126 Bari (Italy), blonda@iesi.ba.cnr.it
‡IMGA-CNR, Via Gobetti 101, 40129 Bologna (Italy), parmi2@imga.bo.cnr.it
§IESI-CNR, satalino@iesi.ba.cnr.it

# 1 Introduction

In recent years, the scientific community has been involved in an important debate on the reasons why the many image processing techniques presented in the literature have had such a slight impact on their potential field of application [1], [2], [3]. To take this debate into account, we have attempted to make this paper as "focused on its very original core" [1] as possible. This original core is regarded as a set of useful items that a broad scientific audience would be willing to include in commercial image-processing all-purpose software toolboxes [1].

Picture segmentation is a subjective and context-dependent cognitive process. In mathematical terms, this process is ill-posed, i.e., there is no single goal for picture partition algorithms [4], [5]. To make picture segmentation a well-posed problem, it is regarded as an optimization task featuring a firm statistic foundation. Recently, there has been considerable interest in statistical clustering techniques for image segmentation inspired by the methods of statistical physics, which were developed to study the equilibrium properties of large, lattice-based systems consisting of identical, interacting components [6]. In a clustering technique for image segmentation, each pixel is associated with one of a finite number of categories (also termed pure substances, or colors, or region types, or states) to form disjointed regions.

It is well known that, when applied to image segmentation, the traditional hard $c$-means clustering algorithm features one main limitation: it assumes that each region type is characterized by uniform intensity throughout the image, i.e., it assesses global intensity parameters that do not account for local intensity variations. As a consequence, the $c$-means clustering algorithm tends to generate noisy (salt-and-pepper) segmentation results [7]. The same problem affects any noncontextual (pixel-wise) classification algorithm (independent model, [10]) by assuming the independence of observables $y$ given pixel labels (states) $x$ (see (2) below) as well as the independence of pixel labels, such as the Maximum Likelihood (ML) parameter estimation procedure [9], [10], [11]. It is obvious that adjacent pixels are likely to have similar labels when the pixel's size is smaller than the smallest detail of interest. Therefore, to obtain smooth segmentations, a Markov Random Field (MRF) model (for a general overview of MRF models, refer to [12]) is often imposed on the spatial distribution of different image categories to introduce spatial correlation in pixel labeling (*interpixel class dependency*) [6], [7], [8], [9], [10], [13], [14]. This underlying MRF model can be considered a "stabilizer" in the sense of the regularization theory [10], which helps in solving otherwise ill-posed problems [15], [16]. Parameters of the MRF model, i.e., the neighborhood size and the smoothing parameters (clique potentials), incorporate *a priori* knowledge into the modeling process, i.e., they are either user-defined or known on an *a priori* basis, e.g., by means of off-line [6], [8], [9], [21], or on-line [9] parameter estimation techniques employing supervised data. The MRF parameters affect the size of disjointed regions, i.e., the amount of spatial details preserved by the segmentation process [9]. In [21], a detail-preserving image segmentation method utilizing MRFs is proposed. In this algorithm, minimization of the energy function also requires selection of the most appropriate neighborhood system for the pixel under analysis. In [22] and [23], coarse-to-fine multiresolution segmentation approaches are proposed with no adaptive neighborhood being employed, although in [23] smoothing parameters are fixed as a function of scale. The multiresolution segmentation algorithm

proposed in [23] is found to be less likely to be trapped in local minima than the Iterative Conditional Mode (ICM) algorithm [9], since, at each resolution, regions are classified and used to guide finer resolutions. This conclusion is complementary to the observation that convergence of MRF-based segmentation algorithms can be improved by varying the local neighborhoods [24]. In conclusion, adaptive and multiresolution approaches to parameter estimation in segmentation algorithms exploiting MRF models appear to be highly desirable.

Several clustering algorithms that exploit a MRF model to impose spatial continuity in pixel labeling assume that the intensity of each region type is uniform throughout the image [6], [8], [9], [13], and/or employ hard pixel labeling [6], [7], [8], [9]. When these two conditions hold true, these segmentation algorithms can be considered a generalization of the traditional $c$-means clustering algorithm, such that pixels are hard-clustered on the basis of both their intensity and their context (neighboring labels). From an information-processing perspective, it is well known that a hard-decision approach to parameter estimation is less effective than the use of soft decisions which, in picture segmentation applications, would compare the probabilities of all possible pixel labels [10]. This consideration led to the development of fuzzy clustering algorithms for image classification [17], in which the Expectation-Maximization (EM) procedure [18] can be employed to determine the Maximum Likelihood (ML) estimate of parameters of a Gaussian mixture decomposition [19], where a MRF model is often imposed on the spatial distribution of the pixel labels [10], [13].

An interesting algorithm that includes multiresolution analysis and adaptive spatial constraints in a traditional hard $c$-means clustering algorithm is the Pappas Adaptive Clustering (PAC) algorithm for image segmentation [7]. Unlike the $c$-means algorithm, this scheme assumes that each image category is characterized by a slowly varying intensity function, i.e., the same region type may feature different intensity parameters in different parts of the image. PAC is iterative, alternating between estimating category local intensities and hard pixel labels. Moreover, PAC adapts and progressively reduces the neighborhood size: local intensities of each category are estimated by averaging over a sliding window whose size decreases monotonically as the algorithm approaches convergence. This means that PAC starts with global parameter estimates and progressively adapts these estimates to the image's local properties. Parameter estimates in each iteration employ current hard pixel labels, assuming they are correct. Exploitation of a decreasing window size should mitigate deficiencies in crisp label assignments [10]. Finally, PAC can employ a multiresolution implementation scheme. Since PAC preserves large image features while removing small details, it is employed to provide sketches or caricatures of the original image.

This paper proposes a Modified version of PAC (MPAC) to be employed in cascade with a hard noncontextual $c$-means clustering stage. MPAC aims to: i) enhance pattern-preserving capability of both $c$-means and PAC algorithms, such that genuine but small regions are not removed; and ii) feature high usability, because no user-defined parameter is required. Experimental results confirm that MPAC may be considered an easy-to-use and effective algorithm for use in cascade with a $c$-means clustering algorithm to segment images featuring smooth surfaces and no texture. Within the frame of the debate started by Zamperoni [1], we recommend inclusion of the proposed algorithm in a commercial image-processing all-purpose software toolbox.

The organization of this paper is as follows: Section 2 presents a brief review of statistical

clustering algorithms for image segmentation; Section 3 reviews the PAC algorithm; in Section 4 the MPAC scheme is proposed. Experimental results are discussed in Section 5; conclusions are presented in Section 6.

## 2 Statistical approaches to image analysis

Let us consider clustering algorithms for image segmentation defined as optimization problems. We identify as $y_i$ the observed vector at pixel $i$, which belongs to a two-dimensional region (image) $S$, such that $i \in \{1, n\}$, where $n$ is the total number of pixels. Each pixel $i$ can take one label (pure substance) $x_i \in \{1, c\}$. An arbitrary labeling (clustering) of $S$ is denoted by $x = (x_1, ..., x_n)$. Some probabilistic methods provide estimate $\hat{x}$ of the true (although unknown) labeled scene $x^*$, which is chosen to have maximum probability, given observables $y$. By Bayes' theorem, $\hat{x}$ maximizes the posterior probability

$$P(x/y) \propto l(y/x)p(x), \tag{1}$$

where $l(y/x)$ is the class conditional density and $p(x)$ is the prior. Thus, $\hat{x}$ is the maximum *a posteriori* (MAP) estimate of $x^*$ that maximizes the posterior cost function (1). A common assumption is that given any particular scene $x$, observables are conditionally independent and each pixel value has the same class conditional density function $f(y_i/x_i)$, dependent only on $x_i$. Then,

$$l(y/x) = \prod_{i=1}^{n} f(y_i/x_i). \tag{2}$$

In the hypothesis that the 2-D random field (stochastic process) $\{p(x)\}$ is any locally dependent Markov Random Field (MRF), then

$$P(x_i/x_{S_{\bar{i}}}) \equiv p_i(x_i/x_{N_i}), \tag{3}$$

where $x_{S_{\bar{i}}}$ is the scene reconstruction anywhere, but pixel $i$, $p_i$ is specific to pixel $i$, and $x_{N_i}$ is the scene reconstruction in the neighborhood of pixel $i$ to be defined according to an $m-$th order MRF (typically, second-order; for more details refer to [6], [8], [20]). Suppose that our goal is to estimate $\hat{x}_i$ at pixel $i$ given all observables $y$ and current reconstruction $\hat{x}_{S_{\bar{i}}}$ elsewhere, i.e., $\hat{x}_i$ maximizes $P(x_i/y, \hat{x}_{S_{\bar{i}}})$. According to assumptions (2) and (3), we can write [9]

$$P(x_i/y, \hat{x}_{S_{\bar{i}}}) \propto f(y_i/x_i)p_i(x_i/\hat{x}_{N_i}). \tag{4}$$

In an iterative optimization procedure the "hats" in Equation (4) imply the use of estimated label assignments from the previous iteration in the current iteration. Since

$$P(x/y) = P(x_i, x_{S_{\bar{i}}}/y) = P(x_i/y, x_{S_{\bar{i}}})P(x_{S_{\bar{i}}}/y),$$

maximization of Equation (4) guarantees that $P(x/y)$ never decreases at any maximization step, i.e., convergence to a local maximum of $P(x/y)$ is assured. When addressing pixel $i$ with Equation (4), only $y_i$, $x_i$ and the labels of the neighbors are required. Implementation of Equation (4) is trivial for any MRF $\{p(x)\}$, which is locally dependent. While maximization of Equation (4) is performed at every point in the image, label updating

can be implemented, for example, in a batch mode (at the end of each raster scan). This iterative suboptimal procedure, termed the Iterated Conditional Modes (ICM) algorithm, guarantees convergence to a local maximum of $P(x/y)$ in few processing cycles (e.g., about six [9]). Then, ICM must be started with a good initialization of the scene $x$, which is often provided by a noncontextual ML classifier [9]. ICM has been criticized for its heavy dependence on initial classification [25], and because it ignores the fuzzy membership (degree of compatibility) with which a pixel is associated with a given class [9], [11], [26]. For example, when applied to satellite image applications, fuzzy classification can be employed to identify misclassified cases and to direct ground surveys [11], [17], [27]. Simulated Annealing (SA) exploiting the "Gibbs sampler" is capable of reaching the global maximum of

$$P_T(x/y) \propto \{l(y/x)p(x)\}^T, \tag{5}$$

where $T > 0$ is a user-defined parameter representing the "absolute temperature" of the system, when at each pixel $i$ label $x_i$ is chosen to maximize the posterior probability [6], [9]

$$P_T(x_i/y, x_{S_{\bar{i}}}) \propto \{f(y_i/x_i)p_i(x_i/x_{N_i})\}^{1/T}. \tag{6}$$

Observe that ICM, which provides simple suboptimal iterative scene estimates, is equivalent to SA when $T$ is fixed to 1 [9].

If pairwise interactions in a second-order MRF $\{p(x)\}$ are considered exclusively (i.e., three- and four-point cliques are ignored), and a single smoothing parameter (two-point clique potential) $\beta$ is employed (i.e., all neighbor pairs are treated equally), then [7], [9], [12],

$$p_i(x_i/\hat{x}_{N_i}) \propto exp[\alpha(x_i) + \beta\hat{u}_i(x_i)], \tag{7}$$

where $\alpha(x_i)$, related to one-point clique potential, is the *a priori* knowledge of the relative likelihood of category assignment $x_i$ [7], while $\hat{u}_i(x_i)$ is the current number of 8-adjacency neighbors of $i$ having label $x_i$. Counter $\hat{u}_i(x_i)$ is termed "self-aura measure" [20]. Complementary to the self-aura measure is the "cross-aura measure", $\hat{v}_i(x_i)$, defined as the current number of 8-adjacency neighbors of $i$ having a label different from $x_i$, such that

$$\hat{u}_i(x_i) + \hat{v}_i(x_i) = 8.$$

Parameter $\beta$ in Equation (7) can be considered an intensive quantity, independent of the number of particles (pixels) present in the system; this term describes pairwise interaction potential between two categories that can be related to pure substances or fluids. To continue with the analogy to pure substances, the self-aura measure, which is an extensive quantity, increases when the spatial clumpiness of the clustered (segmented) image increases, that is, when the separability between cluster types increases, which is to say when the common boundary between different clusters decreases. Vice versa, the cross-aura measure increases when the mixing between pairs of different clusters increases, i.e., when the common boundary between pairs of clusters increases.

As underlined by several authors [7], [28], the MRF model by itself is not very useful, unless we provide a good model for class conditional density $f(y_i/x_i)$. Modeling of joint class conditional density $l(y/x)$ affects MAP estimate $\hat{x}$ not only because the model is simply involved in characterizing the posterior cost function (1), but also because it affects estimate

$\hat{x}$ when used to provide starting scene $x$ (e.g., determined by the use of a noncontextual ML classifier) for a contextual iterative classification method such as ICM [28].

Two major models can be used in formulating class conditional densities. The first model, termed continuous random field model [28], employs the causal autoregressive model, the simultaneous autoregressive model, or the conditional Markov model [28], to describe statistical dependency of a gray level at a lattice point on that of its neighbors, given the underlying classes (*interpixel feature correlation*, [14]), e.g., see [10], [23]. It is computationally expensive and may be preferred in modeling images featuring distinct texture information [10], [28]. It will not be further considered in this paper, since we address images featuring little useful texture information. The second model in formulating class conditional densities is the conventional spectral model based on a multivariate-Normal assumption for the distribution of independent spectral responses [28]: in the hypothesis that each category $j \in \{1, c\}$ has uniform intensity $\mu(j)$ and that the image is corrupted by a white Gaussian noise field independent of the scene and featuring standard deviation $\sigma$, then the conditional density term $f(y_i/x_i)$ becomes [7], [9], [10]

$$f(y_i/x_i) = exp\left\{-\frac{1}{2\sigma^2}[y_i - \mu(x_i)]^2\right\}. \tag{8}$$

This spectral model is computationally simple and yields good classification results when applied to images holding little useful texture information [28]. Substituting Equations (7) and (8) in Equation (4), and setting $\gamma = 1/2\sigma^2$, we obtain the final form of the suboptimal iterative solution to the maximization of (1), which is

$$\hat{x}_i = arg \min_{x_i \in \{1,c\}} \{\gamma[y_i - \mu(x_i)]^2 - \alpha(x_i) + \beta\hat{v}_i(x_i)\}, \quad \forall i \in S. \tag{9}$$

The first term on the right side of Equation (9), called the *error term* [15], constrains intensity of the region type $\mu(x_i)$ to be close to data $y_i$, i.e., it represents fidelity of the region type to the data [9]. The third term on the right side of Equation (9) imposes spatial continuity in pixel labeling to reduce "mixing" between different pure substances (categories). In the framework of regularization theory, the sum $[-\alpha(x_i) + \beta\hat{v}_i(x_i)]$ is termed *smoothness functional* or *stabilizer*. A stabilizer embodies our *a priori* knowledge of the labeled scene and is used to impose constraints on the solution, since significant probabilities are assigned only to underlying scenes that satisfy these constraints [15]. According to regularization theory, parameter $\beta$ in Equation (9) is a *regularization parameter* that determines the tradeoff between the strength of the *a priori* assumptions about the solution and the closeness of the solution to the data. For example, as noise (degree of uncertainty) increases, then $\beta$ should increase, to find a solution diminishingly closed to the raw data.

It is to be observed that $\gamma$ and $\beta$ are inversely correlated: reducing $\gamma$ (increasing $\sigma$) is equivalent to increasing $\beta$ to obtain smoother segment contours while larger details are lost, and vice versa [7]. When $\beta = 0$, Equation (9) is equivalent to an ML classifier ignoring contextual information. In [9], it is remarked that the exact value of $\beta$ is usually unimportant with ICM if smaller values are used on earlier iterations [9], so that at an early stage the algorithm follows the data, while at later stages the algorithm follows the region model [7]. However, in many practical applications, $\beta$ is kept constant [7], [9], [23] because the schedule needed for changing this parameter would require additional free parameters

which are generally image dependent [23]. If the 2-D stochastic process {p(x)} is to support special features such as thin lines, then pairwise interaction models do not suffice, and smoothing parameters $\beta$s, as well as the neighborhood size, must be adapted depending on the type of scene and, possibly, on local image properties [8], [9], [21].

## 3 Pappas clustering for image segmentation

In [7], Equation (9) is adapted to exploit both types of contextual information, namely, interpixel feature correlation given the underlying classes and interpixel class dependency [14] (see Sections 1 and 2). Equation (9) is modified as follows

$$\hat{x}_i = arg \min_{x_i \in \{1,c\}} \{\gamma[y_i - \hat{\mu}_{W_i}(x_i)]^2 - \alpha(x_i) + \beta \hat{v}_i(x_i)\}, \quad \forall i \in S, \tag{10}$$

where the model for conditional density (error term) employs a slowly varying intensity function $\hat{\mu}_{W_i}(x_i)$ estimated as the average of the gray levels of all pixels that currently belong to region type $x_i$ and fall inside a window $W_i$ which is centered on pixel $i$. The window size decreases monotonically as the algorithm approaches convergence to guarantee robust estimation of intensity functions as the segmentation, which is "crude" in the early stages of the algorithm, becomes progressively more sensitive to local image properties [7]. When the number of pixels of type $x_i$ within window $W_i$ is less than or equal to window width $W_{i,w}$, then estimate $\hat{\mu}_{W_i}(x_i)$ is not considered reliable and pixel $i$ cannot be assigned to region type $x_i$. Thus, isolated regions with area smaller than $W_{i,w}$ are removed by the clustering algorithm. As shown in [7], the new spectral model of the error term proposed in Equation (10) makes PAC more robust than traditional $c$-means clustering algorithms in the choice of the number of clusters [19], because regions of entirely different intensities can belong to the same category, as long as they are separated in space.

In Equation (10), Pappas also fixes $\alpha(x_i) = 0, \forall x_i \in \{1, c\}$, i.e., all pixel states are assumed to be equally likely, and employs a constant $\beta = 0.5$ for all iterations and for any image. Vice versa, coefficient $\gamma$, which is inversely related to noise variance, is either user-defined or assessed from the image of interest to control the amount of details detected by the algorithm (which increases as $\gamma$ increases). Then, Equation (10) becomes

$$\hat{x}_i = arg \min_{x_i \in \{1,c\}} \{\gamma[y_i - \hat{\mu}_{W_i}(x_i)]^2 + 0.5 \cdot \hat{v}_i(x_i)\}, \quad \forall i \in S, \tag{11}$$

where $\gamma$ is a free parameter.

In [7], a PAC hierarchical multiresolution implementation is proposed to reduce the amount of computation. This implementation constructs a pyramid of images at different resolutions by low pass filtering and decimating by a factor of two. At each level in the pyramid, the algorithm uses the segmentation generated at the previous level, expanded by a factor of two, as a starting point. The $\gamma$ parameter doubles when the resolution level decreases (i.e., noise standard deviation is reduced by half, then algorithm follows the data at low resolution stages), while parameter $\beta$ is kept unchanged (otherwise, $\gamma$ can be fixed while $\beta$ is reduced by half). This multiresolution approach, besides reducing computation time, may improve feature-preserving capability (for details, refer to [7]). The Pappas adaptive clustering algorithm for image segmentation at a given resolution level is shown

in Fig. 1. For implementation details about how to reduce computation time in estimating the intensity parameters, refer to [7].

# 4 Modification of Pappas' algorithm

Our main objective is to segment noiseless images featuring smooth surfaces and no texture while preserving genuine but small details. Real data sets satisfy these pictorial constraints. For example, high resolution satellite images acquired by SPOT HRV or Landsat TM sensors may be not severely degraded by noise. If this is the case, the model used to express the joint class-conditional distribution can be based on a multivariate-Normal assumption of independent spectral responses [28], which is consistent with Equations (9) and (10). Therefore, to reach our goal, a feasible approach is to develop a Modified PAC (MPAC) algorithm featuring enhanced pattern-preserving capability. Thus, in line with the debate started by Zamperoni [1], we do not claim to present yet another picture segmentation algorithm. As an adaptation of Equation (10), we adopt the following cost function

$$\hat{x}_i = arg \min_{x_i \in \{1,c\}} \{\triangle(x_i)\}, \quad \forall i \in S, \tag{12}$$

such that

$$\triangle(x_i) = \begin{cases} \min\{[y_i - \hat{\mu}_{W_i}(x_i)]^2, [y_i - \hat{\mu}_W(x_i)]^2\}, & \\ \text{if } \hat{\mu}_{W_i}(x_i) \text{ exists and is considered reliable;} & (13) \\ [y_i - \hat{\mu}_W(x_i)]^2, \text{ if } \hat{\mu}_{W_i}(x_i) \text{ does not exist} & \\ \text{or is considered unreliable,} & (14) \end{cases}$$

where $W$ identifies the fixed window that covers the entire image.

A scheme of the MPAC algorithm at a given resolution level is shown in Fig. 2. The main differences between MPAC and PAC, emerging from the comparison of Figs. 1 and 2 and Equations (10) to (12) respectively, can be summarized as follows:

1. In MPAC, since $\beta = 0$ (see Equations (10) and (12)), no spatial continuity in pixel labeling is enforced, i.e., contextual information is employed to model class conditional distributions exclusively. Since it ignores any regularization term in its objective function, MPAC focuses on keeping the error term low, so that local intensity parameter $\hat{\mu}_{W_i}(x_i)$ of region type $x_i$ is kept as close as possible to observable data $y_i$. This also means that MPAC assumes to deal with noiseless raw images (see Section 2). Since it employs no regularization coefficient, MPAC is also easy to use: in cascade to a non-contextual $c$-means clustering algorithm, MPAC requires no user-defined parameter.

2. Rather than alternating estimation of local intensity parameters and pixel labels, MPAC alternates between estimating local intensity parameters, pixel labels, and global intensity parameters.

3. If no local intensity average is available to model one category in the neighborhood of a given pixel, MPAC employs global statistics to determine whether that category is the one closest to pixel data. Note that when an intensity estimate $\hat{\mu}_{W_i}(x_i)$, computed in neighborhood $W_i < W$ centered on pixel $i$, does not exist or is considered unreliable

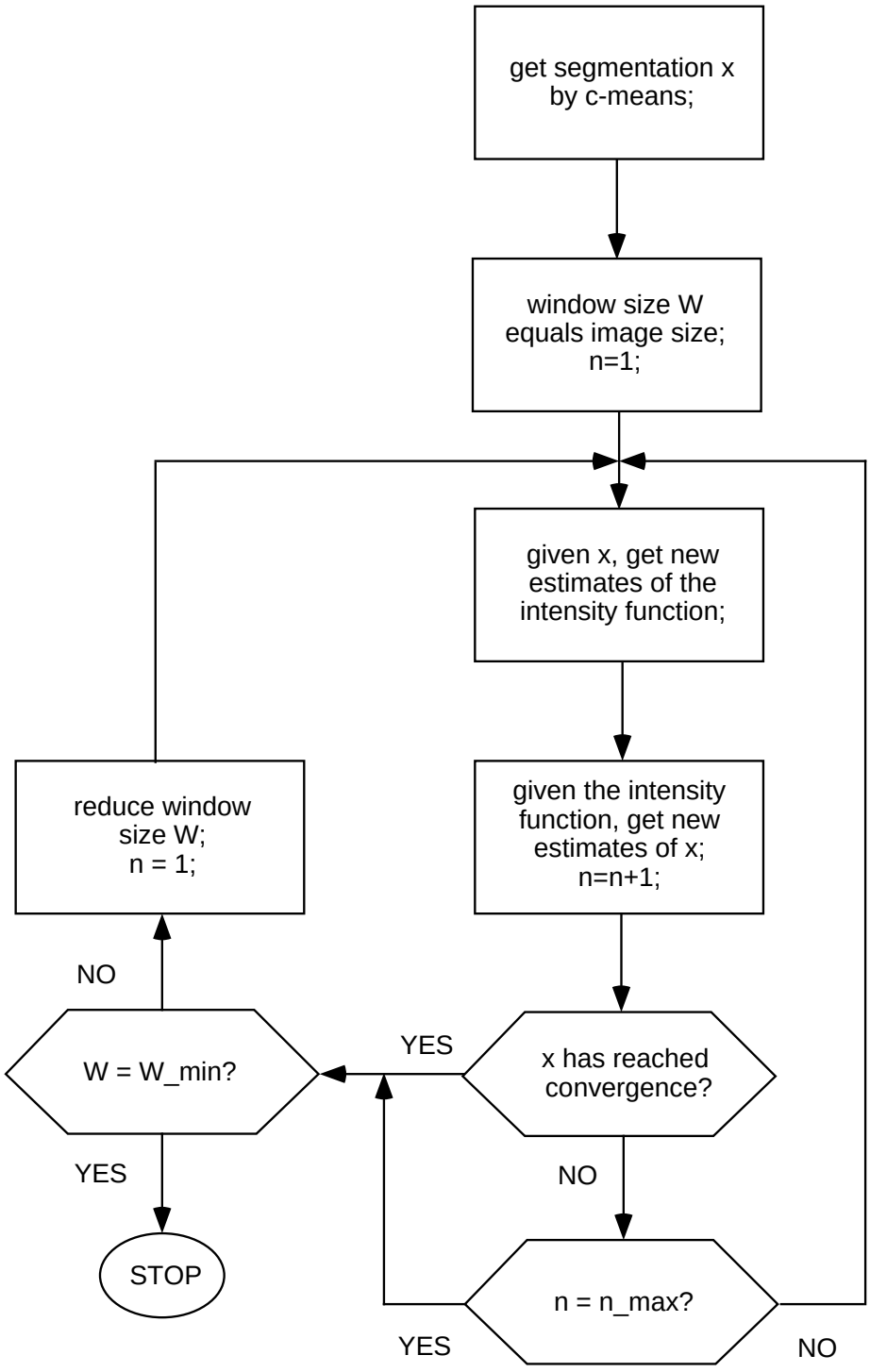


Figure 1: The Pappas adaptive clustering algorithm for image segmentation at a given resolution level.

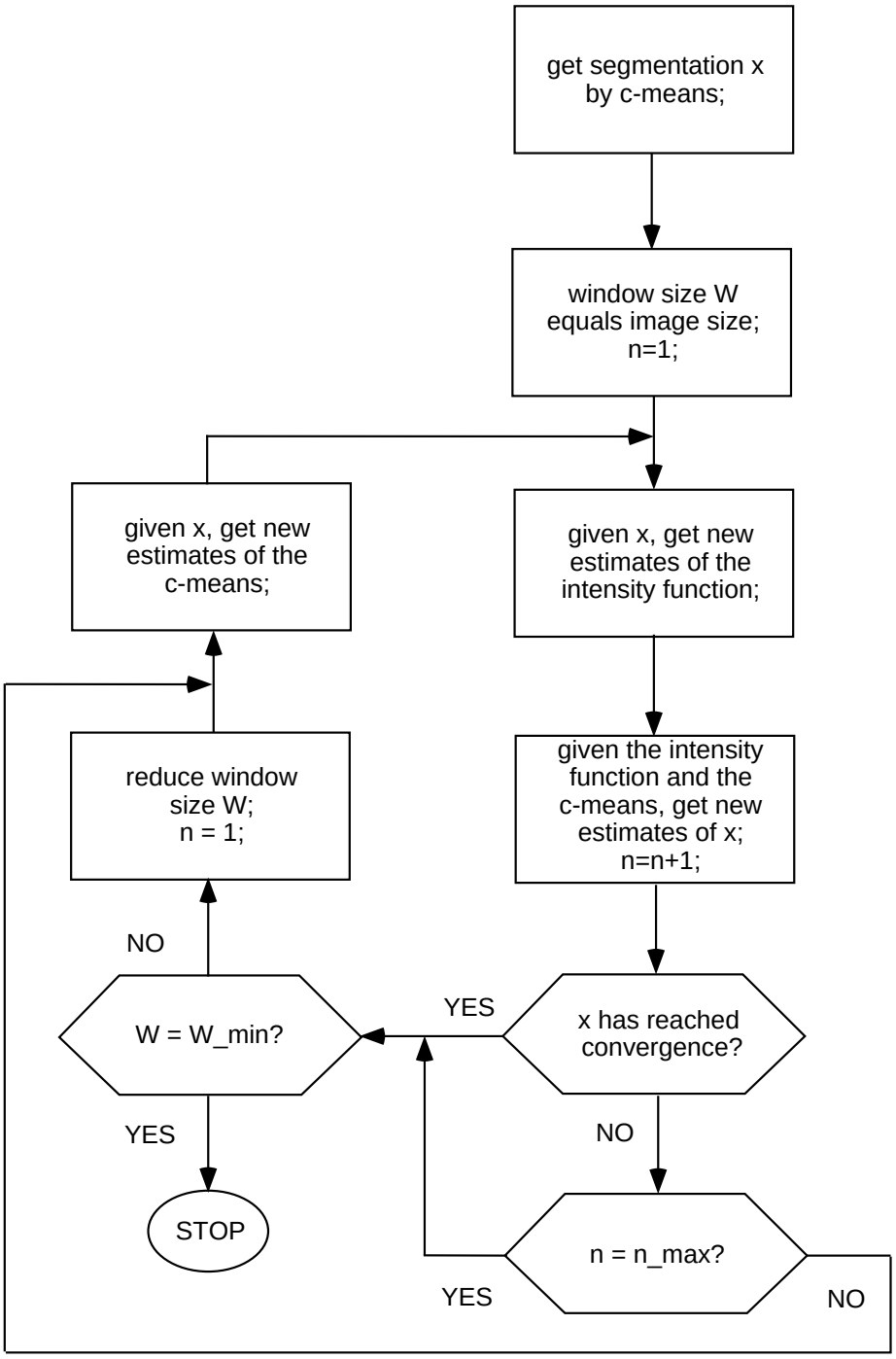


Figure 2: The modified Pappas adaptive clustering algorithm for image segmentation at a given resolution level.

by Equation (12), then global intensity $\hat{\mu}_W(x_i)$, which is the image-wide estimate of the average gray value of all pixels currently belonging to region type $x_i$, is employed instead for comparison with the pixel data according to Equation (14). In line with Equation (10), local estimate $\hat{\mu}_{W_i}(x_i)$ is not considered reliable by Equation (12) when the number of pixels of type $x_i$ within window $W_i$ is less than or equal to adaptive window width $W_{i,w}$. This adaptation of the Pappas algorithm is sufficient to avoid removal of isolated but genuine regions whose area is smaller than $W_{i,w}$.

4. In MPAC, when local intensity estimate $\hat{\mu}_{W_i}(x_i)$ exists and is considered reliable by Equation (12), *both local and global intensity estimates*, $\hat{\mu}_{W_i}(x_i)$ and $\hat{\mu}_W(x_i)$ respectively, are employed for comparison with the pixel data according to Equation (13). It is worthy of mention that we have found images to which Equations (12) to (14) apply successfully while exclusive exploitation of local intensity estimates in Equation (13) is uncapable of providing satisfactory image partitions.

Theoretical failure modes and limitations of the MPAC algorithm are listed below:

1. MPAC can be applied only to images where sharp intensity transitions occur at region boundaries, i.e., where each category is characterized by a slowly varying intensity function. In fact: i) MPAC employs no texture model; and ii) owing to its crisp labeling strategy, MPAC is unable to ignore noisy pixels while segmentation parameters are estimated.

2. MPAC is a suboptimal iterative algorithm detecting local minima. This is tantamount to stating that MPAC largely depends on its initialization. To initialize MPAC, any hard $c$-means clustering algorithm can be chosen from those found in the literature, see for example [30], [31]. One main issue regards the number of clusters to be detected [19], [23]. Nonetheless, in line with PAC, MPAC is more robust than $c$-means clustering algorithms in the choice of the number of clusters, because regions of entirely different intensities can belong to the same category, as long as they are separated in space.

3. MPAC exploits higher degree of heuristics than PAC, i.e., MPAC features a statistical framework which is less rigorous than that featured by PAC.

# 5 Quantitative evaluation of segmentation

Most picture partitions are evaluated visually and qualitatively on a subjective, perceptual basis. This is also due to the fact that traditional supervised measures employed in image processing tasks, such as the mislabeling rate, are global statistics unable to account for local visual properties. For example, the segmentation of a crossboard picture may improve when the number of small holes decreases even though its mislabeling rate actually increases [22]. Therefore, these global statistics cannot be employed for quantitative evaluation of picture segmentation [29]. Since no single segmentation goal exists because of the subjective appraisal of continuous perceptual features, a system developed to compare segmentation results must employ [4], [5]: i) an entire set of measures of success (termed *battery test*) to account for the fuzziness of perceptual segmentation; ii) a *test set of images* to allow

exploitation of supervised data, i.e., of *a priori* knowledge about the objects in a scene, so that the external environment (supervisor) provides the generic (vague) segmentation task with an explicit goal; and iii) a set of exisiting techniques against which the proposed algorithm must be compared.

Let us start by defining the battery test. As we are dealing with image segmentation algorithms that minimize a given cost function, i.e., algorithms that are well-posed and feature one explicitly defined segmentation objective, this function is included in the battery test of the system we have developed to compare segmentation results. Other statistics that are considered in the segmentation comparison are: average value of the error term, number of cycles required to reach convergence, number of label replacements, and number of misclassified pixels when ground truth data are available. Note that computation time is not considered in the battery test since the proposed MPAC algorithm may employ a PAC-based hierarchical multiresolution implementation that significantly reduces the amount of computation (for details, refer to [7]).

The test set of images must consist of a sufficient number of real and standard data sets capable of demonstrating the potential utility of MPAC, i.e., these images must feature a high signal-to-noise ratio and little texture. We selected a standard achromatic human face (Lenna), a three-band SPOT HRV image, and a Landsat TM image provided with supervised data fields. Because of difficulties in comparing alternative classification procedures in a meaningful way, the first application uses a face as opposed to natural scenes because we know what a face looks like and can therefore judge the results of the clustering algorithm intuitively [11].

To demonstrate its potential utility, MPAC is compared against two existing segmentation algorithms based on probabilistic theory. The first of these clustering algorithms is the PAC iterative procedure implemented to detect a local minimum of the following cost function:

$$\hat{x}_i = arg \min_{x_i \in \{1,c\}} \{[y_i - \hat{\mu}_{W_i}(x_i)]^2 + \beta \cdot \hat{v}_i(x_i)\}, \quad \forall i \in S, \tag{15}$$

where $\beta$, rather than $\gamma$, as in Equation (11), is the free parameter to be user-defined, which makes the error term in Equation (15) more similar to that employed in Equations (13) and (14). Experimentally, we discovered that in our test images several pixels may feature no reliable intensity estimate in their adaptive neighborhood at some resolution level. Thus, we had to remove the Pappas constraint that considers any estimate $\hat{\mu}_{W_i}(x_i)$ unreliable when the number of pixels of type $x_i$ in window $W_i$ is less than or equal to window width $W_{i,w}$. In other words, in our implementation of the Pappas algorithm, any category $x_i$ featuring pixel occurrence above zero within window $W_i$ is considered in the minimization of Equation (15).

The second segmentation scheme employed for comparison is the SA algorithm, which is a general purpose approach for detecting the absolute minimum of a cost function [32]. We employ SA to minimize the cost function:

$$\hat{x}_i = arg \min_{x_i \in \{1,c\}} \{[y_i - \mu(x_i)]^2 + \beta \cdot \hat{v}_i(x_i)\}, \quad \forall i \in S, \tag{16}$$

where $\beta$ is the free parameter to be user-defined, and $\mu(x_i)$ is the non-adaptive (image-wise) vector parameter of region type $x_i$ to be provided by a noncontextual classifier (e.g., the hard

$c$-means clustering algorithm). Note that Equation (16) is a special form of Equation (9) to be compared with Equation (15). According to the principles of simulated annealing [8], as temperature $T$ decreases, the chance of random label assignments decreases accordingly. At each iteration, $T$ is lowered by a constant cooling rate $\delta$, which is fixed at 0.95. Pixels are visited according to a raster scan and updated in batch (synchronous) mode at the end of each cycle [9]. Unfortunately, SA requires long computation times to run successfully.

### 5.1 Standard image application

The standard achromatic input image of Lenna is shown in Fig. 3. Six category templates are fixed by a photointerpreter: $\mu(1) = 68$ (corresponding to surface classes: shadow wall, hair), $\mu(2) = 100$ (hair), $\mu(3) = 143$ (wall, hat), $\mu(4) = 160$ (skin), $\mu(5) = 190$ (hat), $\mu(6) = 213$ (wall, skin). This set of templates is larger than what is suggested in [7], where 2 to 4 clusters were employed to obtain caricatures of the original images. These templates are employed by the noncontextual hard $c$-means clustering algorithm to provide the three segmentation algorithms under testing with an initial segmentation to start from. To highlight functional differences between the three algorithms, free parameter $\beta$ in Equations (15) and (16) is kept low. This choice yields SA and PAC segmentation results where small spatial details tend to be preserved, and can be compared to segmentation results of MPAC, which implicitly adopts parameter $\beta = 0$.

Fig. 4 shows the output of the Simulated Annealing (SA) algorithm exploiting parameters $T = 800$, $\delta = 0.95$, $\beta = 0.5$, $t_{max} = 150$, where $t_{max}$ is the epoch at which the algorithm is stopped [32]. Since weight $\beta$ of the term enforcing spatial continuity in pixel labeling is small, then, as expected, minimization of Equation (16) provides a segmentation result almost identical to that generated by the hard $c$-means clustering algorithm at the initialization step. For this reason the initial segmentation provided by the hard $c$-means clustering is not shown. Figs. 5 to 7 show the plots of the mean value of Equation (16), the mean value of the error term in Equation (16), and the percentage number of replacements per pixel (multiplied by factor 100) respectively. All plots are in terms of the epoch number. Of course, since $\beta$ is small, the two plots in Figs. 5 and 6 respectively are quite similar.

Fig. 8 shows the output of the PAC algorithm exploiting parameter $\beta = 0.5$, $t_{max} = 150$, where $t_{max}$ is the epoch at which the algorithm is stopped. Figs. 9 to 11 provides meaningful plots of this PAC application at full resolution. PAC reaches convergence after about 10 iterations, although the cost function tends to oscillate once its asymptote has been reached. Since weight $\beta$ of the term enforcing spatial continuity in pixel labeling is small, then minimization of Equation (15) is mostly focused on minimization of the error term. This is shown in Figs. 9 and 10 where plots of the PAC mean cost and error term appear similar. Observe that although $\beta$ is the same as the one in the SA application, small details are better preserved in Fig. 8 than in Fig. 4 (e.g., in the woman's hat). The conclusion is that the enhanced feature-preserving capability of PAC with respect to SA is totally dependent on their different class conditional density models (error term). As expected, Figs. 9 to 11 also show that PAC reaches convergence earlier than the time-consuming SA algorithm at full resolution.

Fig. 12 shows the output of the MPAC algorithm exploiting parameter $t_{max} = 150$, where $t_{max}$ is the epoch at which the algorithm is stopped. Since MPAC does not enforce

Figure 3: Standard achromatic input image of Lenna (521x512 pixels).

Figure 4: Output of the SA algorithm applied to Fig. 3. SA parameters are: $T = 800$, $\delta = 0.95$, $\beta = 0.5$, $t_{max} = 150$.

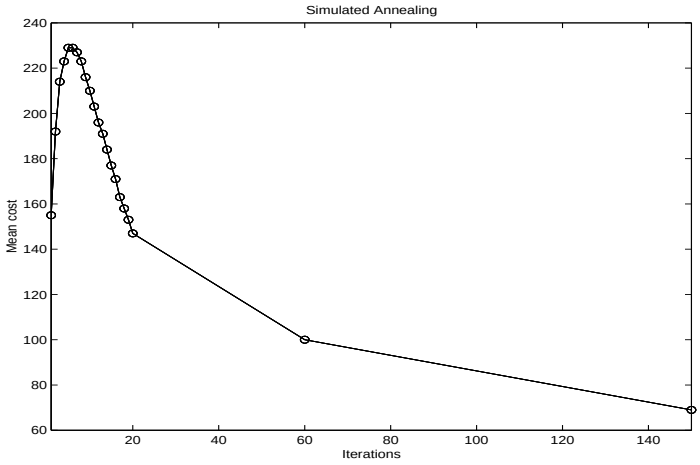


Figure 5: Plot of the SA algorithm applied to Fig. 3: mean value of Equation (16).

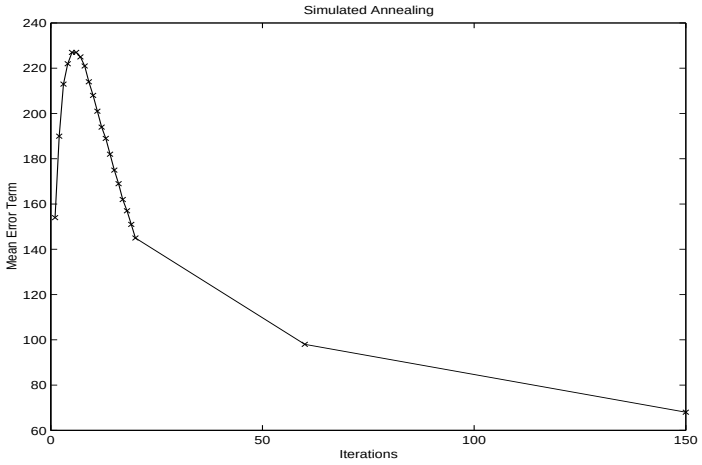


Figure 6: Plot of the SA algorithm applied to Fig. 3: mean value of the error term in Equation (16).

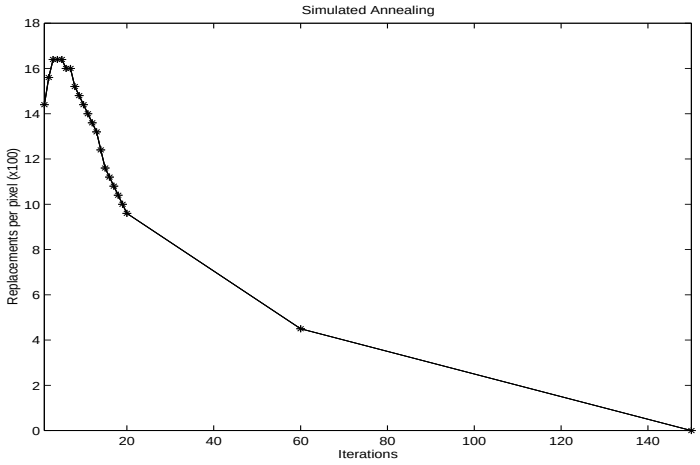


Figure 7: Plot of the SA algorithm applied to Fig. 3: the percentage number of replacements per pixel.

Figure 8: Output of the PAC algorithm applied to Fig. 3. PAC parameters are: $\beta = 0.5$, $t_{max} = 150$.

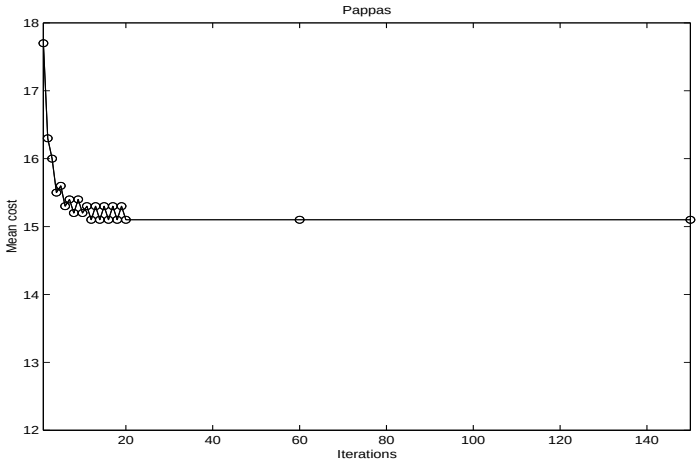


Figure 9: Plot of the PAC algorithm applied to Fig. 3: mean value of Equation (15).

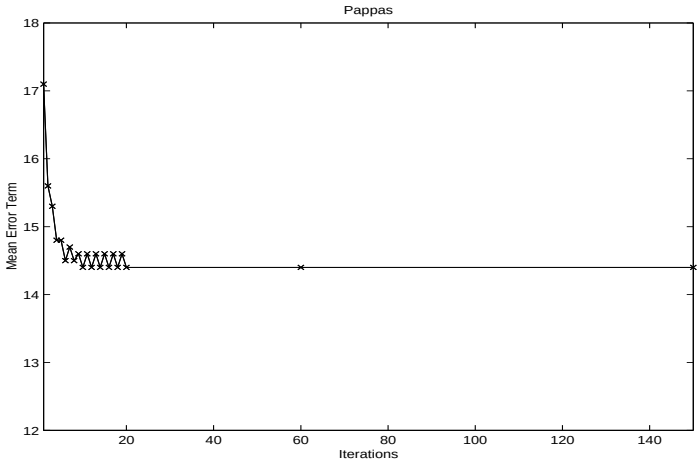


Figure 10: Plot of the PAC algorithm applied to Fig. 3: mean value of the error term in Equation (15).

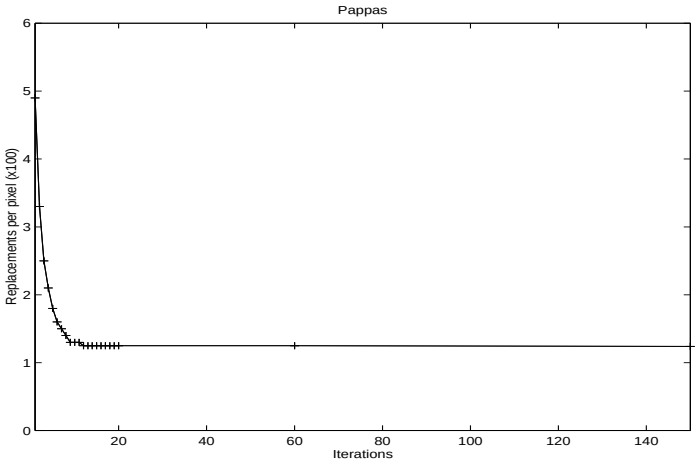


Figure 11: Plot of the PAC algorithm applied to Fig. 3: percentage number of replacements per pixel.

any spatial continuity in pixel labeling, small details are preserved better in Fig. 12 than in Fig. 8 (e.g., in the woman's hat). MPAC tolerates isolated pixels, which are considered a problem in [7] even though they feature high contrast to their neighbors. This creates a sort of dithering effect around the woman's shoulder, for example. Overall, subjective appraisal of Fig. 12 with respect to Fig. 8 may be either positive or negative, accounting for the subjective nature of segmentation problems. However, this subjective assessment may not be relevant, with Fig. 12 representing a limiting product (when $\beta = 0$) of MPAC, which is a (slightly modified, see Section 4) PAC algorithm capable of generating, in its traditional form, Fig. 8 when $\beta$ is small, starting from Fig. 3. Figs. 13 and 14 provide meaningful plots of this MPAC application at full resolution. MPAC reaches convergence after about 20 iterations. To be compared with initial centroid values, final templates estimated by the MPAC algorithm are: $\mu(1) = 69.1$, $\mu(2) = 107.1$, $\mu(3) = 138.8$, $\mu(4) = 161$, $\mu(5) = 185.4$, $\mu(6) = 206$.

### 5.2 Unsupervised satellite image application

Figs. 15 to 17 show a multispectral SPOT HRV image of the city of Porto Alegre (Rio Grande do Sul, Brazil) acquired on Nov. 7, 1987. Spectral bands are Green, Red and Near InfraRed, respectively. We underline the presence of three bay bridges linked to the large island in the upper left-hand corner of these images. A zoomed area around the city airport extracted from Fig. 17 is shown in Fig. 18.

Eight category templates are fixed by a photointerpreter: $\mu(1) = (49, 43, 18)$ (water), $\mu(2) = (39, 30, 66)$ (vegetation), $\mu(3) = (46, 30, 112)$ (vegetation), $\mu(4) = (49, 42, 36)$ (airport, asphalt), $\mu(5) = (66, 66, 85)$ (street), $\mu(6) = (207, 193, 152)$ (metal roofs), $\mu(7) = (50, 46, 62)$ (houses), $\mu(8) = (57, 51, 62)$ (houses). This set of templates is larger than that suggested in [7] to obtain caricatures of the original images. These templates are employed by the noncontextual hard $c$-means clustering algorithm to provide PAC and MPAC with an initial segmentation to start from. This initial segmentation is shown in Fig. 19. The zoomed area taken from Fig. 19 and corresponding to Fig. 18 is shown in Fig. 20. This clustered image shows the presence of several isolated pixels which are typical of noncontextual classification.

To highlight functional differences between the three algorithms, free parameter $\beta$ in Equations (15) and (16) is kept rather low. This choice yields SA and PAC segmentation results in which small spatial details tend to be preserved (i.e., the algorithm follows the data, rather than following the prior region model [7]), and can be compared to MPAC segmentation performances. As in [7], the standard deviation of noise in each spectral band of the satellite image was assumed to be $\sigma = 4$ gray levels, then $\gamma = 1/2\sigma^2 = 0.031$. In this case, from Equation (11), ratio $\beta/\gamma = 0.5/0.031$ is approximately equal to 16. In Equations (15) and (16), this parameter condition is equivalent to fixing $\beta = 16$ given $\gamma = 1$.

Fig. 21 shows the output of the SA algorithm exploiting parameters $T = 800$, $\delta = 0.95$, $\beta = 16$, and $t_{max} = 150$. Fig. 22 shows the zoomed area extracted from Fig. 21 and corresponding to Fig. 18. As expected, since weight $\beta$ of the term enforcing spatial continuity in pixel labeling is small with respect to the error term, which tends to increase with the number of image bands, minimization of Equation (16) provides segmentation results that are not so different from those depicted in Fig. 19, although spatial details are

Figure 12: Output of the MPAC algorithm applied to Fig. 3. MPAC parameter is: $t_{max} = 150$.

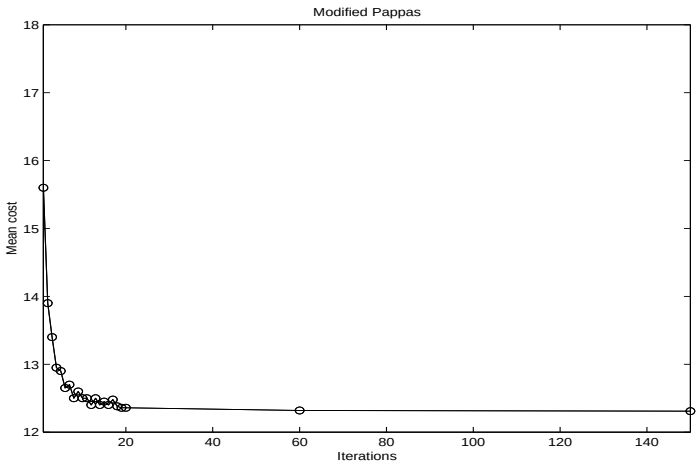


Figure 13: Plot of the MPAC algorithm applied to Fig. 3: mean value of Equation (12)

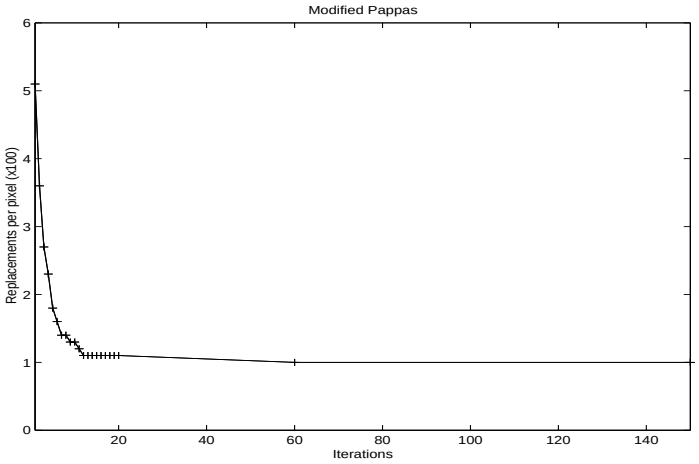


Figure 14: Plot of the MPAC algorithm applied to Fig. 3: percentage number of replacements per pixel.

Figure 15: SPOT HRV image of the city of Porto Alegre (Rio Grande do Sul, Brazil): Band 1 (Visible Green; 512x512 pixels).

Figure 16: SPOT HRV image of the city of Porto Alegre: Band 2 (Visible Red).

Figure 17: SPOT HRV image of the city of Porto Alegre: Band 3 (Near IR).

Figure 18: A zoomed area around the city airport extracted from Fig. 17.

lost (e.g., see the bay bridges in Fig. 21 and some linear patterns in Fig. 22).

Fig. 23 shows the output of the PAC algorithm exploiting parameter $\beta = 16$, and $t_{max} = 150$. The asymptote of the cost function (16) is reached after about 20 iterations at full resolution. Fig. 24 shows the zoomed area extracted from Fig. 23 and corresponding to Fig. 18. Note that although $\beta$ is the same as the one in the SA application, small details are preserved better in Fig. 24 than in Fig. 22. In Fig. 24, the number of isolated pixels is reduced with respect to Fig. 20 (some of them are still present due to oscillations in the estimation of the labeled scene). The conclusion is that the enhanced feature-preserving capability of PAC with respect to SA is totally due to their different class conditional density models (error term).

Fig. 25 shows the output of the MPAC algorithm exploiting parameter $t_{max} = 150$. The asymptote of the cost function (15) is reached after about 15 iterations at full resolution. Fig. 26 shows the zoomed area extracted from Fig. 25 and corresponding to Fig. 18. In Fig. 26, the number of isolated pixels is reduced with respect to Fig. 20. Many isolated pixels, rather than being filtered out, as occurred in Fig. 24, have been linked to neighboring pixels featuring similar spectral signatures. Since MPAC does not enforce any spatial continuity in pixel labeling, small details are better preserved in Figs. 25 and 26 than in Figs. 19 to 24 (e.g., in Fig. 25 the three bay bridges have been reconstructed). To be compared with initial centroid values, final templates estimated by the MPAC algorithm are: $\mu(1) = (49.2, 43.3, 20)$, $\mu(2) = (41.4, 31.6, 73.1)$, $\mu(3) = (45.1, 33.6, 96.4)$, $\mu(4) = (49.2, 44.1, 45.3)$, $\mu(5) = (70, 68.5, 78)$, $\mu(6) = (148.6, 137.2, 105.1)$, $\mu(7) = (50.8, 45.8, 65.4)$, $\mu(8) = (59.6, 55.9, 63.8)$.

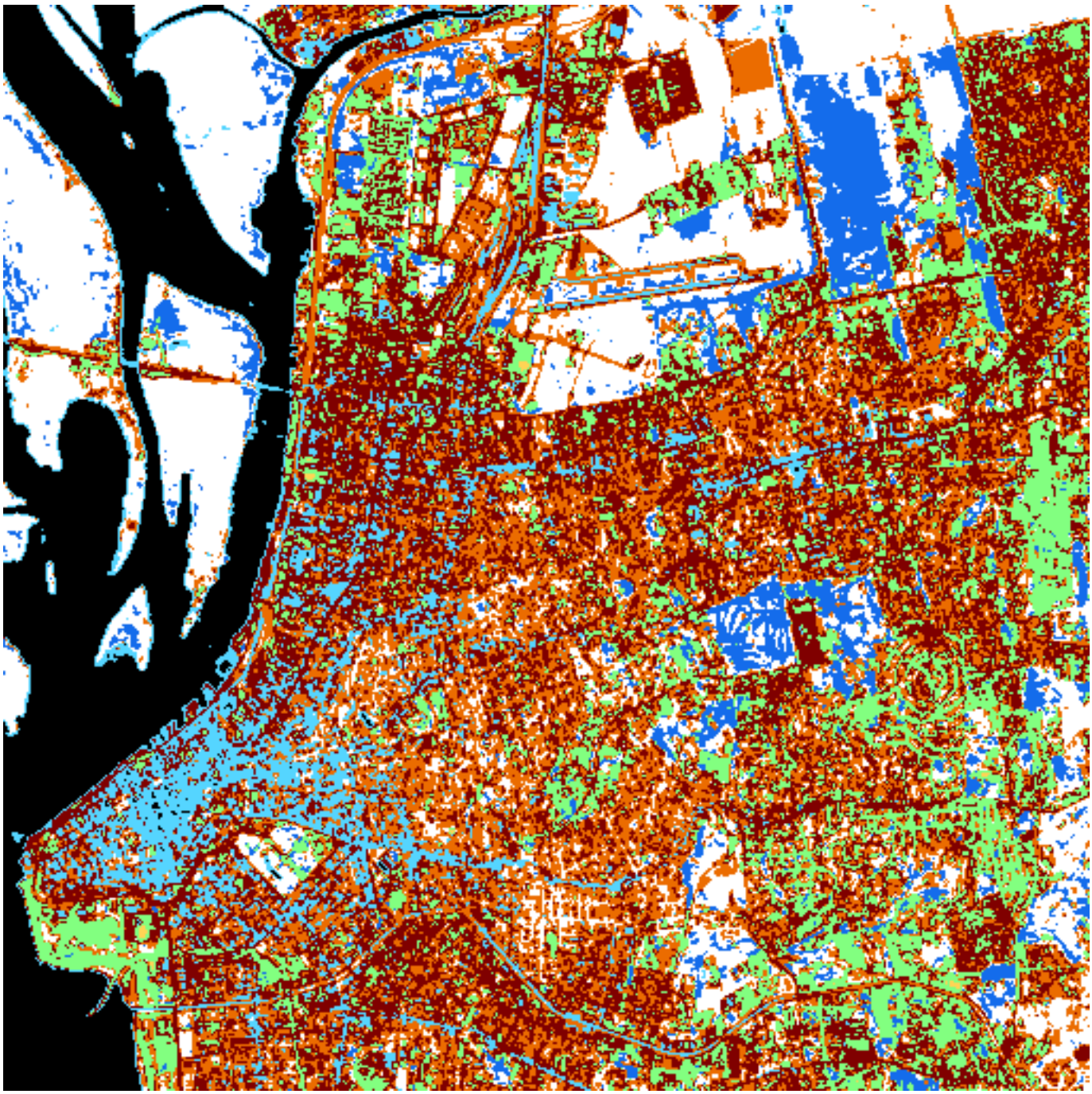

Figure 19: Initial segmentation of Figs. 15-17, obtained by a noncontextual $c$-means clustering algorithm.

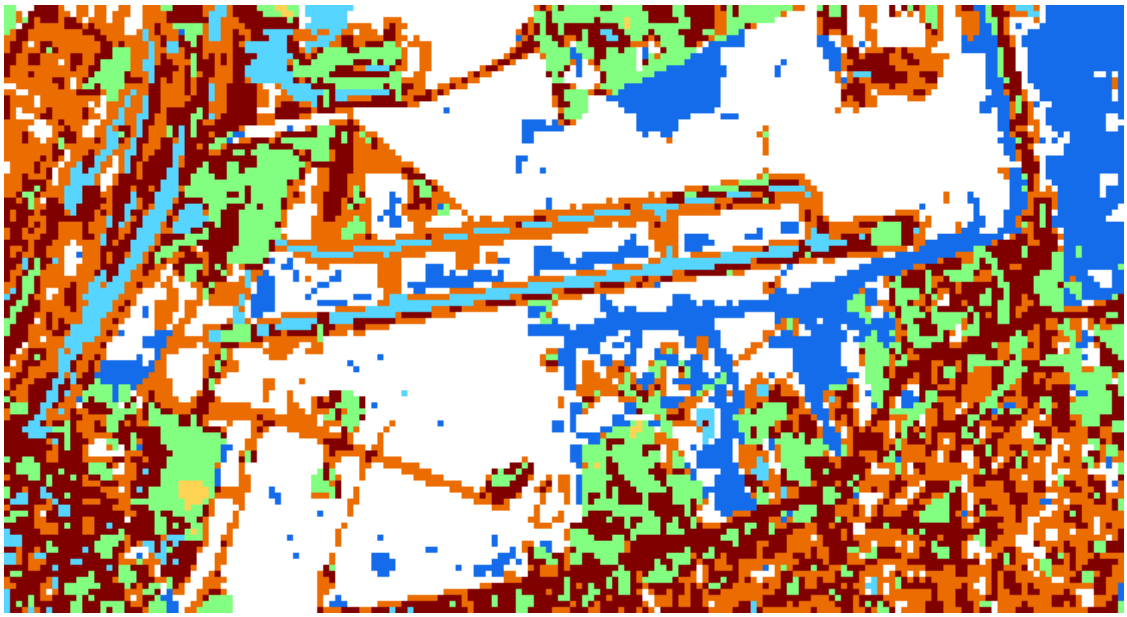

Figure 20: Zoomed area taken from Fig. 19 and corresponding to Fig. 18.

### 5.3 Supervised medical image application

A three-band Magnetic Resonance Image (MRI) of a horizontal section of a brain is shown in Figs. 27 to 29 (Band 1: T1 Magnetization-Prepared RApid Gradient Echo, MP-RAGE; Band 2: T2 Spin-Echo, SE; Band 3: Proton Density, PD). Regions belonging to six classes (tissues) of interest are manually selected by expert photointerpreters. These classes are (see Fig. 30): white matter (red), grey matter (green), Cerebral Spinal Fluid (CSF, liquor: light blue), lesions (yellow), background (blue), other (bones, fat, thalamus: white).

A Self-Organizing Map (SOM, [33]) is employed to extract twenty-two statistical regularities (equivalent to centers of image categories or clusters) from the MRI 3-D histogram. Then each cluster is related to one supervised class by majority voting. All categories relating to the same class form a so-called metacategory, i.e., six metacategories are formed (one for each class). Next, cluster centers detected by SOM are used as input by MPAC, and each image category recomputed by MPAC is related to one supervised class by majority voting. Six new metacategories are formed according to MPAC. Tables 1 and 2 show that, owing to exploitation of contextual information, MPAC improves the average classification performance of the non-contextual (pixel-based) clustering algorithm. Analogous results have been obtained when eleven classes of interest are selected in the MRI image (MPAC average classification performance scoring 64.3 % versus 61.5 % of SOM), and when satellite images provided with ground truth regions are classified.

## 6 Conclusions

PAC is a traditional hard clustering algorithm for image segmentation designed as an optimization task. It is based on a statistical framework where neighborhood adaptivity and

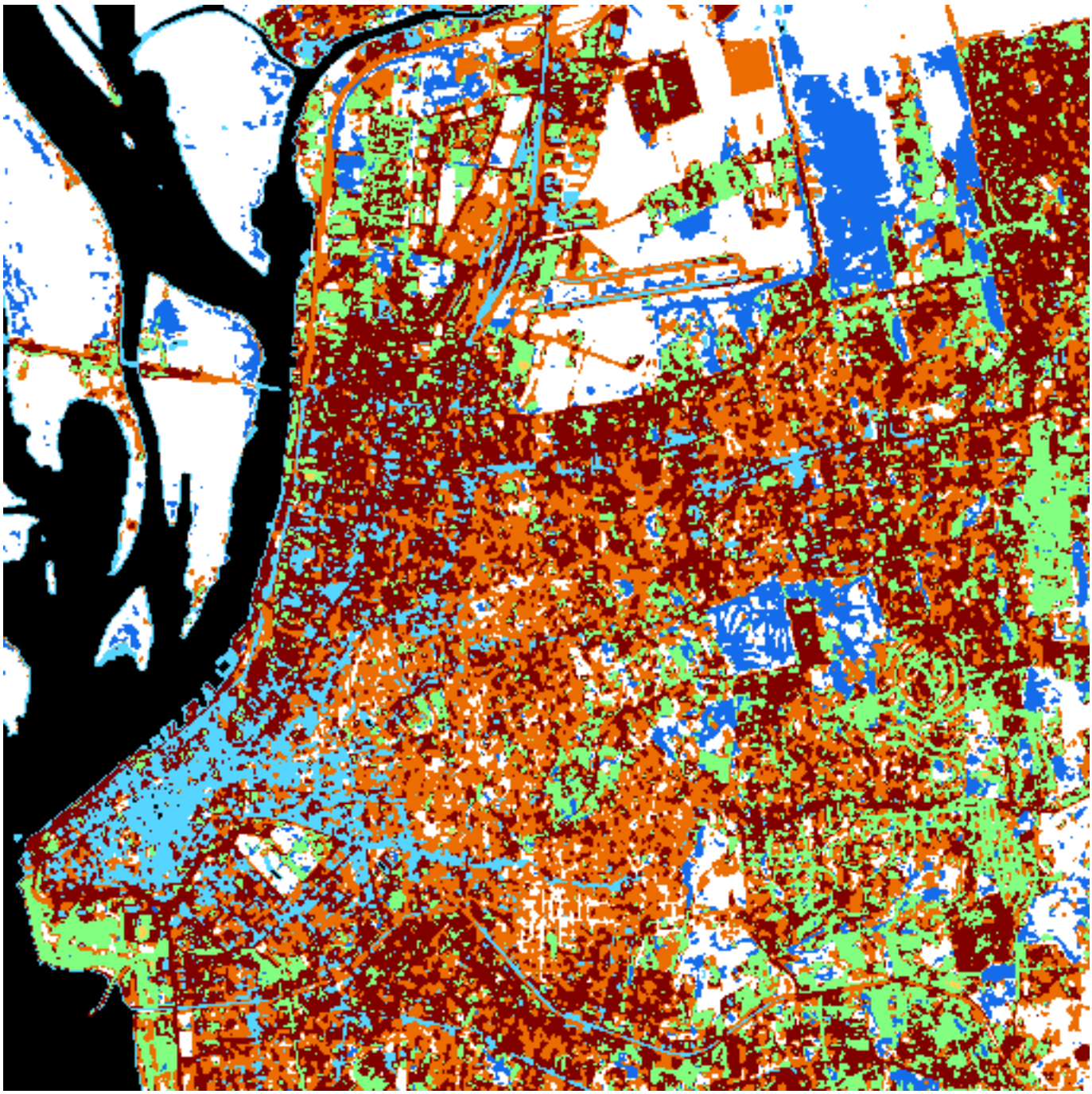

Figure 21: Output of the SA algorithm applied to Figs. 15-17. SA parameters are: $T = 800$, $\delta = 0.95$, $\beta = 16$, and $t_{max} = 150$.

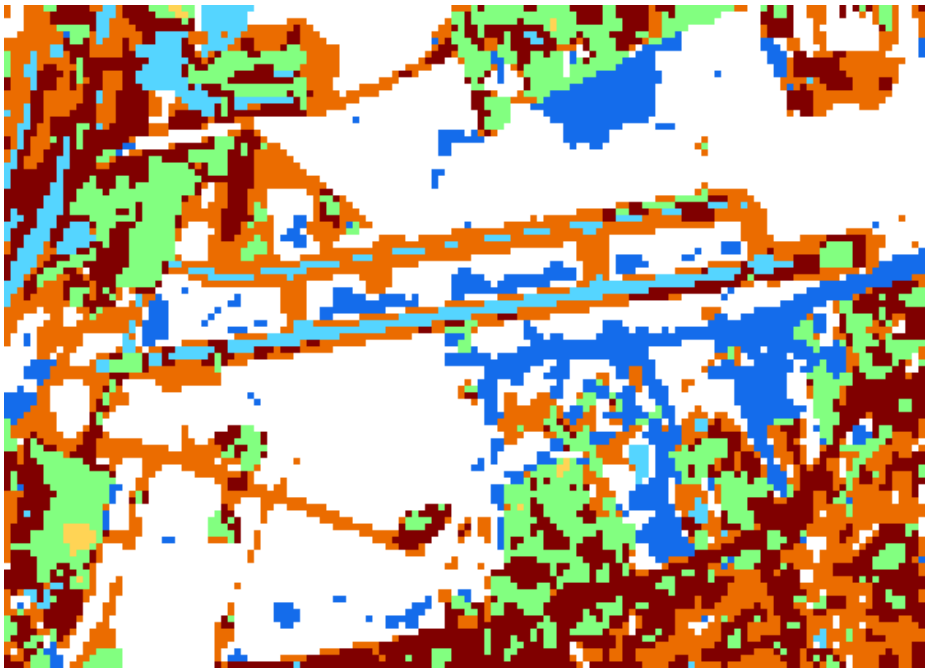

Figure 22: Zoomed area taken from Fig. 21 and corresponding to Fig. 18.

multiresolution analysis are pursued. In this paper, the MPAC algorithm is proposed as a modified version of PAC to feature enhanced pattern-preserving capability with noiseless and textureless images, i.e., when image categories feature slowly varying intensities. These basic assumptions, although severe, reasonably approximate characteristics of several real-world images. MPAC is an iterative suboptimal segmentation algorithm featuring: i) adaptive- and shrinking-neighborhood approach to the estimation of reliable category parameters; ii) a multiresolution approach to improve computation time and segmentation accuracy [7]; iii) hard (crisp) pixel labeling; iv) a spectral model of the error term that accounts for the interpixel feature correlation given the underlying classes (i.e., it exploits contextual information); and v) no interpixel class correlation model of the prior term, i.e., no contextual information is exploited to detect known stochastic components of the labeled scene.

To segment an image featuring no texture and noise, MPAC must be used in cascade with a noncontextual hard $c$-means clustering algorithm (e.g., see [30], [31]), that provides histogram analysis of pixel values. By alternating between estimating pixel labels and local and global intensity values, MPAC preserves small details better than noncontextual hard $c$-means clustering algorithms. Moreover, in line with PAC, MPAC is more robust than $c$-means clustering algorithms in the choice of the number of clusters, because regions of entirely different intensities may belong to the same category as long as they are separated in space [7]. Since MPAC is also easy to use, requiring no user-defined parameter, its exploitation is recommended in a commercial image-processing all-purpose software toolbox [1]: a) to improve segmentation performances of a noncontextual hard $c$-means clustering algorithm; and/or b) to provide initial conditions to hard iterative contextual segmentation

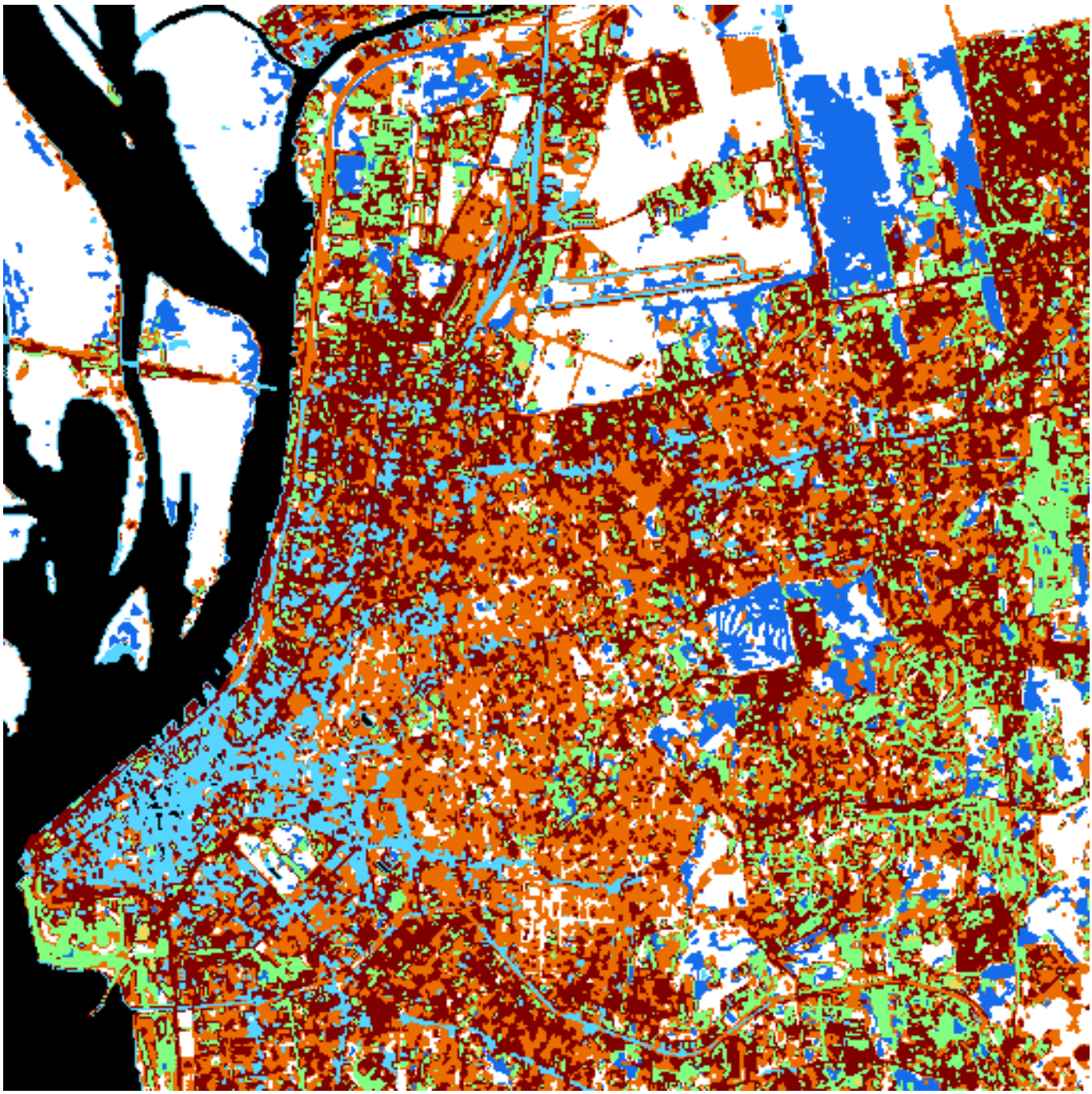

Figure 23: Output of the PAC algorithm applied to Figs. 15-17. PAC parameters are: $\beta = 16$, $t_{max} = 150$.

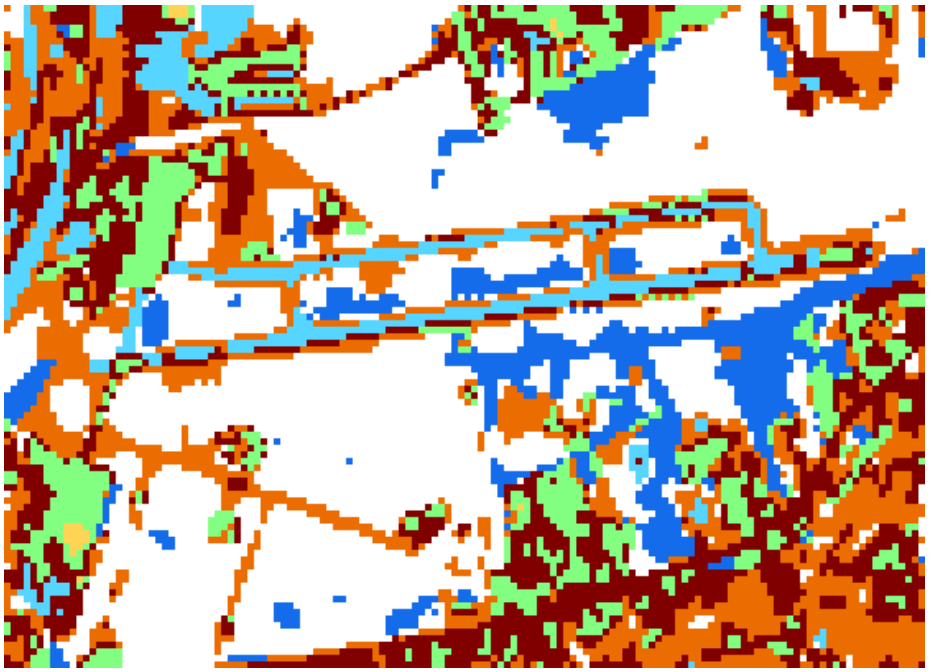

Figure 24: Zoomed area extracted from Fig. 23 and corresponding to Fig. 18.

algorithms where spatial continuity in pixel labeling should be enforced as a monotone increasing function of processing time (i.e., as the algorithm approaches convergence), such as ICM or PAC (see Sections 2 and 3).

Future developments in the field of image segmentation algorithms based on iterative (suboptimal) optimization approaches should employ in parallel:
1) soft decision strategies in pixel labeling to identify pure pixels, mixed pixels and misclassified cases [11], [13], [27], this information being necessary in map accuracy assessment and/or for directing ground surveys [11];
2) adaptive neighborhood approaches, allowing reliable estimation of local pictorial parameters when the iterative procedure alternates between estimates of pixel labels and category parameters [7];
3) multiresolution approaches, to improve computation time and performances [7], [22], [23];
4) locally adaptive combinations of the two types of contextual information, namely, interpixel feature correlation, given the underlying classes, and interpixel class dependency, where all the *a priori* knowledge, if any, is employed to capture (detect) known stochastic components of the labeled scene [7], [9], [21], [28].

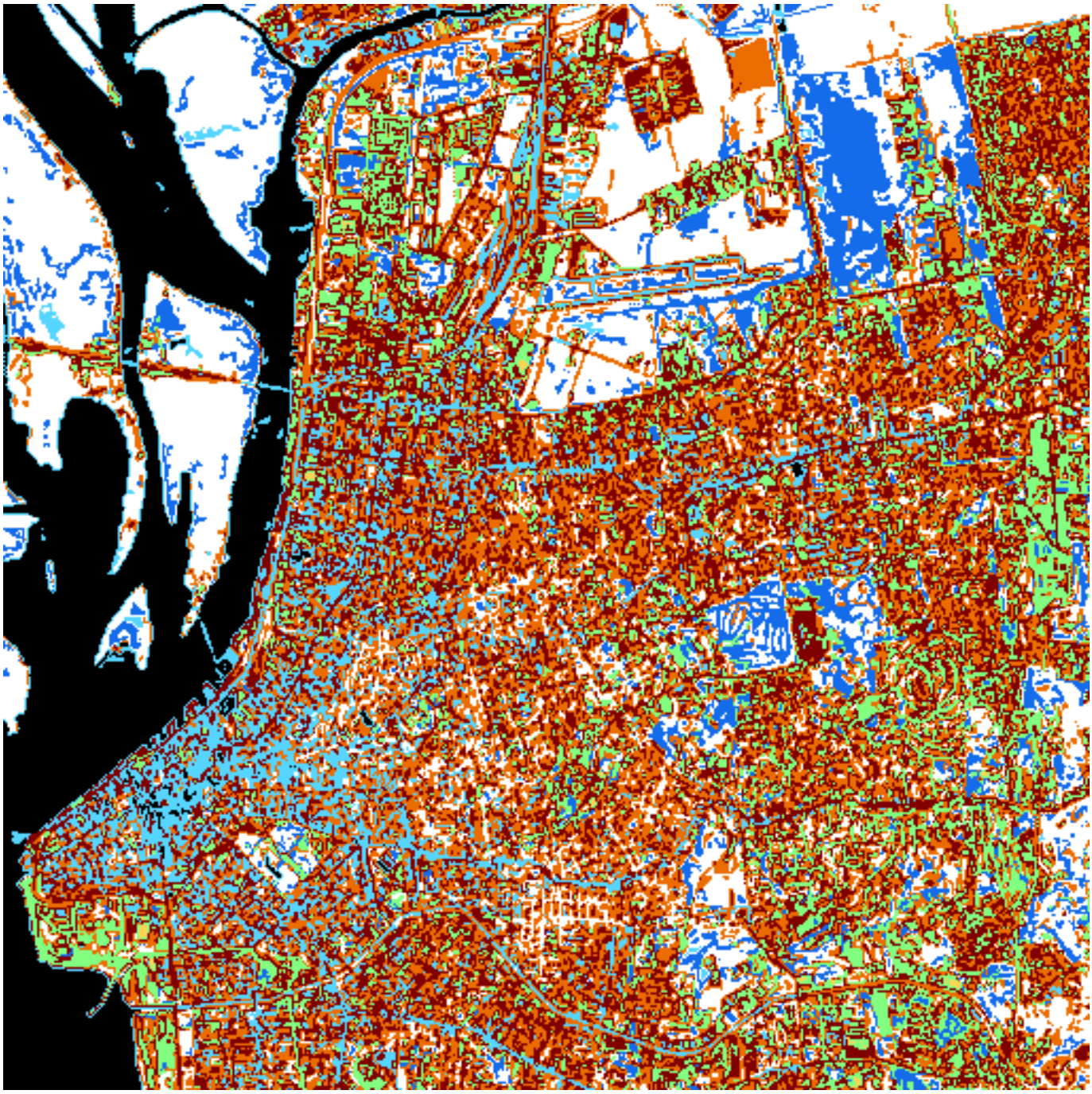

Figure 25: Output of the MPAC algorithm applied to Figs. 15-17. MPAC parameter is: $t_{max} = 150$.

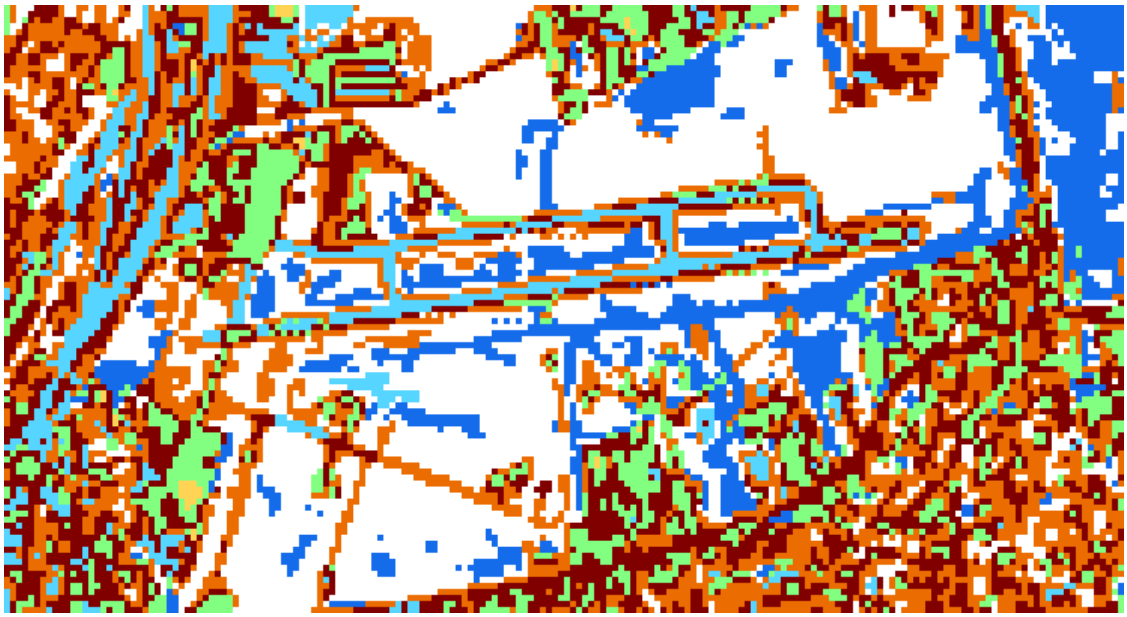

Figure 26: Zoomed area extracted from Fig. 25 and corresponding to Fig. 18.

## References


[1] P. Zamperoni, "Plus ça va, moins ça va," *Pattern Recognition Letters*, vol. 17, no. 7, pp. 671-677, 1996.

[2] R. C. Jain and T. O. Binford, "Ignorance, myopia and naivetè in computer vision systems," *Comput., Vision, Graphics, Image Processing: Image Understanding*, vol. 53, pp. 112-117, 1991.

[3] M. Kunt, Comments on "Dialogue", a series of articles generated by the paper entitled "Ignorance, myopia and naivetè in computer vision systems," *Comput., Vision, Graphics, Image Processing: Image Understanding*, vol. 54, pp. 428-429, 1991.

[4] L. Delves, R. Wilkinson, C. Oliver and R. White, "Comparing the performance of SAR image segmentation algorithms," *Int. J. Remote Sensing*, vol. 13, no. 11, pp. 2121-2149, 1992.

[5] A. Hoover, G. Jean-Baptiste, X. Jiang, P. J. Flynn, H. Bunke, D. G. Goldgof, K. Bowyer, D. W. Eggert, A. Fitzgibbon, and R. B. Fisher, "An experimental comparison of range image segmentation algorithms," *IEEE Trans. Patt. Anal. Machine Intelligence*, vol. 18, no. 7, pp. 673-688, 1996.

[6] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Patt. Anal. Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721-741, 1984.

[7] T. N. Pappas, "An adaptive clustering algorithm for image segmentation," *IEEE Trans. on Signal Processing*, vol. 40, no. 4, pp. 901-914, 1992.

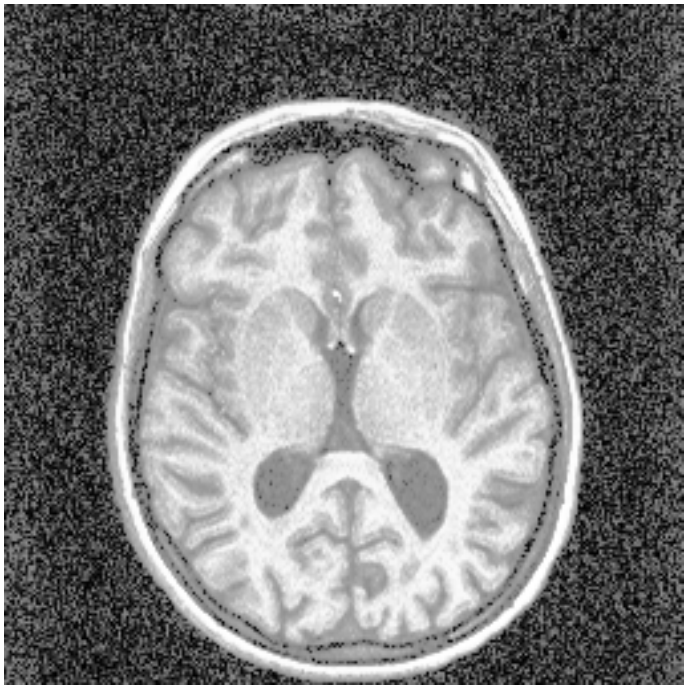

Figure 27: MRI, Band 1: T1 MP-RAGE.

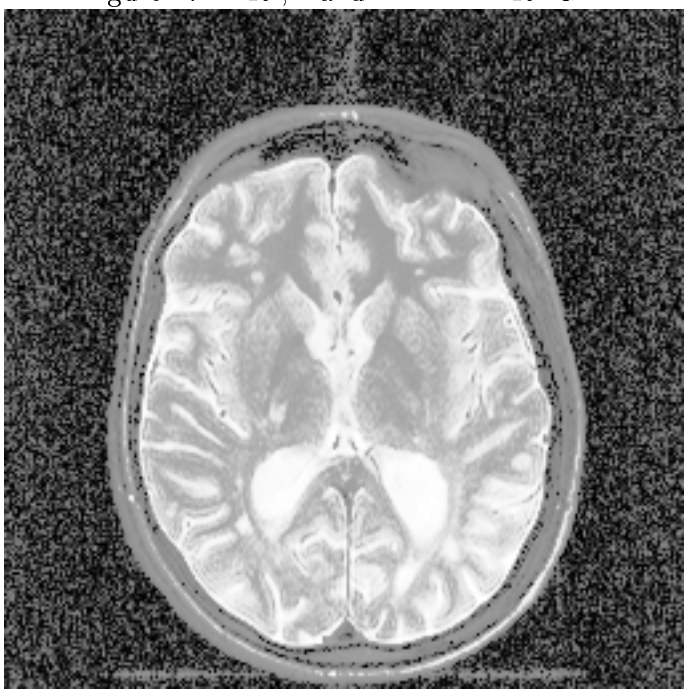

Figure 28: MRI, Band 2: T2 SE.

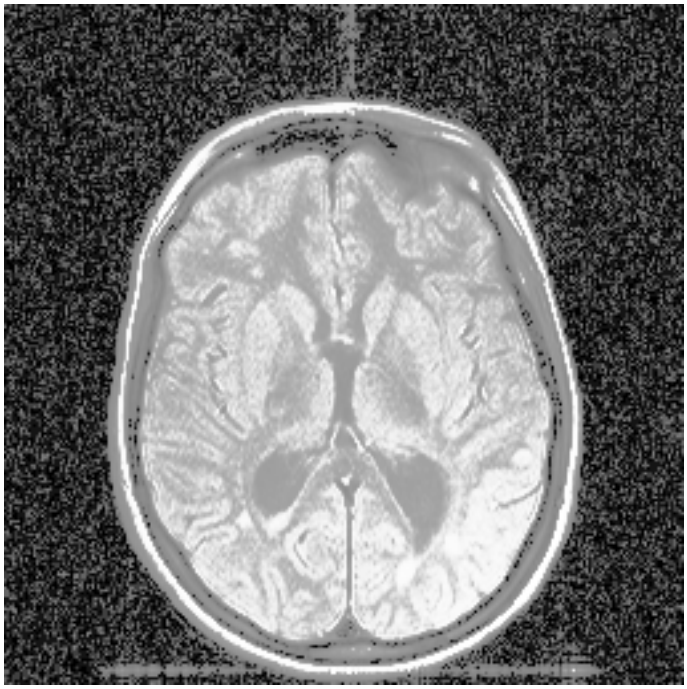

Figure 29: MRI, Band 3: PD.

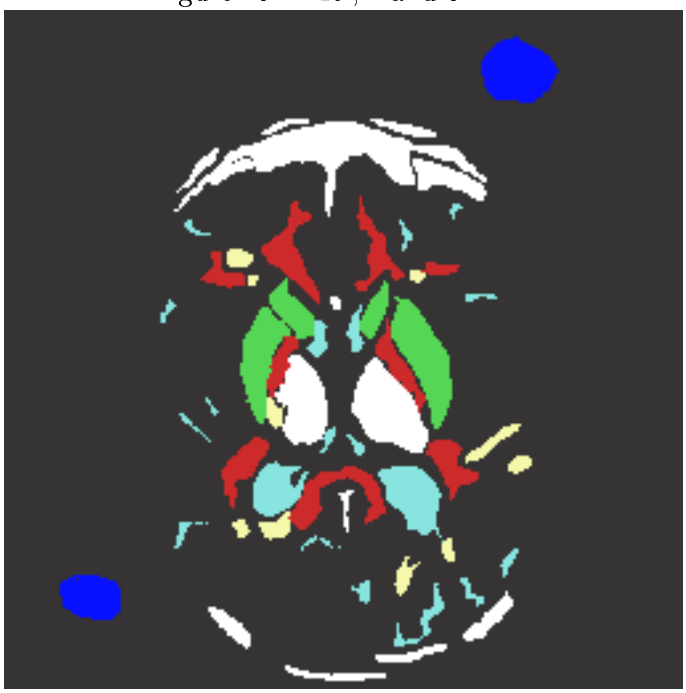

Figure 30: MRI, Areas of interest.

| | Input Class | | | | | | Tot. | Purity |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | Pixels | (%) |
| Metacategory 1 | 1232 | 164 | 4 | 119 | 0 | 505 | 2024 | 60.9 |
| Metacategory 2 | 26 | 723 | 28 | 90 | 0 | 250 | 1117 | 64.7 |
| Metacategory 3 | 0 | 10 | 1168 | 89 | 0 | 74 | 1341 | 87.1 |
| Metacategory 4 | 5 | 2 | 32 | 157 | 0 | 1 | 197 | 79.7 |
| Metacategory 5 | 0 | 0 | 0 | 0 | 727 | 322 | 1049 | 69.3 |
| Metacategory 6 | 412 | 395 | 19 | 24 | 121 | 1928 | 2899 | 66.5 |
| Tot. Pixels | 1675 | 1294 | 1251 | 479 | 848 | 3080 | 8627 | |
| Efficiency (%) | 73.6 | 55.9 | 93.4 | 32.8 | 85.7 | 62.6 | | 68.8 |

Table 1: Confusion matrix: SOM.

| | Input Class | | | | | | Tot. | Purity |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | Pixels | (%) |
| Metacategory 1 | 1213 | 60 | 2 | 40 | 0 | 374 | 1689 | 71.8 |
| Metacategory 2 | 23 | 737 | 35 | 91 | 0 | 243 | 1129 | 65.3 |
| Metacategory 3 | 0 | 16 | 1168 | 85 | 0 | 77 | 1346 | 86.8 |
| Metacategory 4 | 5 | 1 | 24 | 160 | 0 | 1 | 191 | 83.8 |
| Metacategory 5 | 0 | 0 | 0 | 0 | 770 | 275 | 1045 | 73.7 |
| Metacategory 6 | 434 | 480 | 22 | 103 | 78 | 2110 | 3227 | 65.4 |
| Tot. Pixels | 1675 | 1294 | 1251 | 479 | 848 | 3080 | 8627 | |
| Efficiency (%) | 72.4 | 57.0 | 93.4 | 33.4 | 90.8 | 68.5 | | 71.4 |

Table 2: Confusion matrix: MPAC.

[8] H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Trans. Patt. Anal. Machine Intelligence*, vol. PAMI-9, no. 1, pp. 39-55, 1987.

[9] J. Besag, "On the statistical analysis of dirty pictures," *J. R. Statist. Soc. B*, vol. 48, no. 3, pp. 259-302, 1986.

[10] J. Zhang, J. W. Modestino and D. A. Langan, "Maximum likelihood parameter estimation for unsupervised stochastic model-based image segmentation," *IEEE Trans. Image Processing*, vol. 3, no. 4, pp. 404-419, 1994.

[11] P. C. Van Deusen, "Modified highest confidence first classification," *PE&RS*, vol. 61, no. 4, pp. 419-425, 1995.

[12] R. Chellappa and A. Jain, Eds., *Markov Random Fields: Theory and Application.* New York: Academic, 1993.

[13] J. Dehemedhki, M. F. Deami and P. M. Mather, "An adaptive stochastic approach for soft segmentation of remotely sensed images," in *Series in Remote Sensing: vol. 1* (Proc.

of the Int. Workshop on Soft Computing in Remote Sensing Data Analysis, Milan, Italy, Dec. 1995), E. Binaghi, P. A. Brivio and A. Rampini, Eds., World Scientific, Singapore, pp. 211-221, 1996.

[14] A. H. Schistad Solberg, T. Taxt, and A. K. Jain, "A Markov Random Field Model for classification of multisource satellite imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 34, no. 1, pp. 100-113, 1996.

[15] F. Girosi, M. Jones and T. Poggio, "Regularization theory and neural networks architectures," *Neural Computation*, vol. 7, 1995, pp. 219-269.

[16] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, vol. 317, pp. 314-319, 1985.

[17] F. Wang, "Fuzzy supervised classification of remote sensing images," *IEEE Trans. Geosci. Remote Sensing*, vol. 28, no. 2, pp. 194-210, 1990.

[18] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Statist. Soc. Ser. B*, vol. 39, pp. 1-38, 1977.

[19] S. Medasani and R. Krishnapuram, "Determination of the number of components in Gaussian mixtures using agglomerative clustering," *Proc. Int. Conf. on Neural Networks '97*, Houston, TX, June 1997, pp. 1412-1417.

[20] I. Elfadel and R. Picard, "Gibbs random fields, cooccurrences, and texture modeling," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, no. 1, pp. 24-37, 1994.

[21] P. C. Smits, and S. G. Dellepiane, "Synthetic aperture radar image segmentation by a detail preserving Markov Random Field approach," *IEEE Trans. Geosci. Remote Sensing*, vol. 35, no. 4, pp. 844-857, 1997.

[22] J. Liu, and Y. Yang, "Multiresolution color image segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, no. 7, pp. 689-700, 1994.

[23] C. Bouman, and B. Liu, "Multiple resolution segmentation of textured images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 2, pp. 99-113, 1991.

[24] H. Derin, and C. S. Won, "A parallel image segmentation algorithm using relaxation with varying neighborhoods and its mapping to array processors," *Comput. Vision Graphics Image Processing*, vol. 40, pp. 54-78, Oct. 1987.

[25] R. Klein and S. J. Press, "Adaptive Bayesian classification of spatial data," *Journal of the American Statistical Association*, vol. 87, no. 419, pp. 844-851, 1992.

[26] P. B. Chou and C. M. Brown, "The theory and practice of Bayesian image labeling," *Int. Journal of Computer Vision*, vol. 4, pp. 185-210, 1990.

[27] G. M. Foody, N. A. Campbell, N. M. Trodd and T. F. Wood, "Derivation and applications of probabilistic measures of class membership from the maximum-likelihood classification," *PE&RS*, vol. 58, no. 9, pp. 1335-1341, 1992.

[28] Y. Jhung and P. H. Swain, "Bayesian contextual classification based on modified $M$-estimates and Markov Random Fields," *IEEE Trans. Geosci. Remote Sensing*, vol. 34, no. 1, pp. 67-75, 1996.

[29] B. D. Ripley, "Statistics, images, and pattern recognition," *Canadian Journal of Statistics*, vol. 14, no. 2, pp. 83-111, 1986.

[30] B. Fritzke, "The LBG-U method for vector quantization - an improvement over LBG inspired from neural networks," *Neural Processing Letters*, vol. 5, no. 1, pp. 83-111, 1997.

[31] J. C. Bezdek and N. R. Pal, "Two soft relatives of learning vector quantization," *Neural Networks*, vol. 8, no. 5, pp. 729-743, 1995.

[32] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis and Machine Vision*, Chapman & Hall, London, 1993.

[33] T. Kohonen, *Self-Organizing Maps*, Springer Verlag, Berlin, 1995.