Perceptual Smoothing and Segmentation of Colour Textures

M. Petrou, M. Mirmehdi, and M. Coors

Centre for Vision, Speech and Signal Processing,
Surrey University, Guildford GU2 5XH, England
{M.Petrou,M.Mirmehdi}@ee.surrey.ac.uk

Abstract. An approach for perceptual segmentation of colour image textures is described. A multiscale representation of the texture image, generated by a multiband smoothing algorithm based on human psychophysical measurements of colour appearance is used as the input. Initial segmentation is achieved by applying a clustering algorithm to the image at the coarsest level of smoothing. Using these isolated *core clusters* 3D colour histograms are formed and used for probabilistic assignment of all other pixels to the core clusters to form larger clusters and categorise the rest of the image. The process of setting up colour histograms and probabilistic reassignment of the pixels is then propagated through finer levels of smoothing until a full segmentation is achieved at the highest level of resolution.

1 Introduction

The topics of texture segmentation and colour segmentation have attracted the attention of many researchers. At first sight it might seem trivial to solve the problem of colour texture segmentation, as it may appear that the obvious route would be to combine the knowledge gained from the research in the texture area with that gained in the colour area. However, there is a fundamental property which characterises colour texture and which has just emerged from the research in psychophysics [1]: Human perception of colour depends on the spatial frequency of the colour component. In other words, the perceptual response of the human visual system to a certain part of the electromagnetic spectrum depends on the frequency with which this stimulus is spatially distributed. Thus, colours that appear in a multicolour pattern are perceived differently from colours that form uniform areas (e.g. it has been shown [2] that any coloured pattern with frequency higher than 8 cycles per 1° of visual angle is seen as black). Zhang and Wandell [1] actually proposed a new colour system, called SCIE-Lab, which takes into consideration exactly this property of the visual system.

At least two issues seem relevant here. Firstly, the human vision system is able to extract colour textures as single entities without difficulties, with colour

being a segmentation cue that is an integral part of pre-attentive vision. Secondly, image features which characterise a texture at a certain resolution may be entirely different from image features that characterise the same texture at another resolution. Another important characteristic of the human vision system is that it works as a process, with the analysis of a frame relying on a previous grosser analysis. This is achieved either with the help of peripheral vision followed by foveation to the area of interest, or by increasing the physical proximity of the viewed object. The former approach relies on the mechanism of switching sensors (going from the information obtained by the rods to that obtained by the cones). The latter approach, however, relies on using the same sensor but changing the number of degrees of visual angle it occupies in the field of view, in other words changing the spatial frequency a pattern extents to the eye. Both approaches are characterised by causality: the information flows from the coarse level to the finer level of resolution as the coarser analysis precedes the one performed at the finer level.

The reason we are concerned with the human colour perception is because the criteria by which we judge a segmentation to be good or not are subjective. In the absence of specific application requirement, we expect the segmentation of an image to agree with that performed by our own vision system. Thus, our interest in colour texture perception here concerns the way different textures can be perceived as separate homogeneous regions in the preattentive stage of vision.

Inspired by the above observations, in this paper we are proposing a mechanism of segmenting colour textures, by constructing a *causal*, multiscale *tower* of image versions based on perceptual considerations. The reason we call it "tower" and not a "pyramid" is because we do not perform subsampling and thus we preserve the same number of pixels at all levels. The levels of the tower are constructed with the help of blurring masks put forward by Zhang and Wandell [1], by assuming that the same colour-textured object is seen at 1,2,3,... meters distance. Hence, each coarser version of the image imitates the blurred version the human vision system would have "seen" at the corresponding distance. The analysis of the image starts at the coarsest level and proceeds towards the finest level, just like it would have happened if a person was slowly approaching a distant object. The mechanism with which information is transfered from a coarse to a finer level is probability theory that makes use of causality. We do not advocate that this is actually the mechanism deployed by humans; we use this approach because it is a sound mathematical tool that allows the incorporation of both features and preliminary conclusions that refer to many different levels of analysis.

The originality of the work presented is twofold: While multiresolution pyramids have been proposed and successfully utilised for several tasks, including texture segmentation [3–6], it is the first time that a multiscale/multilevel representation of the image which emulates the human colour perception, and which most significantly, takes into consideration the change in spatial frequency of the perceived pattern, is used for segmentation. Secondly, although the issue of

transfer of information from one level of resolution to the next has been tackled by several researchers, and probabilistic relaxation has been used in multiresolution pyramid representations of data [7–10], it is the first time that a probabilistic relaxation theory, appropriate for operating *across* different levels of scale and exploiting a dictionary of permissible label configurations appropriate for *region labelling*, as opposed to edge or line labelling [10], is developed.

In the next section we shall give a brief literature review of the relevant issues. We must stress that there is very little work that is directly concerned with colour textures. In Section 3 we shall describe the method by which the perceptual tower of images is created. In Section 4, we shall present our probabilistic framework of propagating information in the causal direction across the levels of the tower, and in Section 5, the application of the probabilistic framework to colour texture images will be described. In Section 6, we shall present results of our approach when used to segment several colour texture images. We shall compare them with the results obtained by the recently proposed method in [11]. Our conclusions are presented in Section 7.

2   Literature Survey

The main consideration of *texture perception* in the Computer Vision literature has been with the derivation of descriptive features of the underlying texture. For example, Julesz and Bergen [12] used descriptions such as colour, widths, lengths, and orientations of local features, namely *textons*, to explain differences in artificially generated images. Also, Malik and Perona [13] provided a comparison of their computational model of human texture perception with psychophysical data obtained in texture discrimination experiments, while Tamura et al. [14] approximated computationally some basic textural features such as coarseness, directionality, line-likeness, contrast, roughness and regularity, which correspond to human visual perception. Most of these studies did not consider colour information, and moreover, their features are useful when viewing a scene from a fixed distance only.

Segmentation of texture images is a major field of research in Computer Vision. Textures may be regular or randomly structured, and various structural, statistical and spectral approaches have been proposed towards segmenting them [15–17]. One example of a recent technique is by Jain and Farrokhnia [18] who presented a texture segmentation algorithm focussed on a multi-channel Gabor filtering approach which is believed to characterise the processing of visual information in the early stages of the human visual system. Multi-scale approaches for texture analysis are few and far in-between. Unser and Eden[19] extracted texture energy measures form the image and smoothed the output of the extraction filter bank using Gaussian smoothing at different scales. The features in these multiscale planes are reduced, by diagonalising scatter matrices evaluated at two different spatial resolutions, and thresholded to yield texture segmentation. Matalas et al.[20], used a B-spline transform in order to obtain images at several smoothing levels to calculate vector dispersion and gradient ori-

entation at different scales. A small disparity function is then applied to segment textures. Roan et al. [21] describe a method for classification of textured surfaces viewed at different resolutions, i.e. viewed at different scales or distances, while the image size remains constant. They used greylevel cooccurrence matrices and the Fourier power spectrum of an unknown texture image, taken at one of any several resolutions, to classify it as one of six known textures.

None of the above approaches is concerned with colour textures. On the other hand, there is a vast forum of work on *colour image segmentation* e.g. [22–27,11]. In general, most colour texture representation schemes either use a combination of gray level texture features together with pure colour features, or they derive texture features computed separately in each of the three colour spectral channels. For example, Coleman and Andrews [28] used K-means clustering in each colour band and maximised a cluster fidelity parameter for a more psychovisually acceptable segmented image. Tan and Kittler [29] used eight DCT texture features computed from the intensity image and six colour features derived from the colour histogram of a textured image for classification. Panjwani and Healey[30] presented an unsupervised segmentation technique based on Markov Random Fields which clustered a colour image in the RGB space. Their Markov Random Fields approach made use of the spatial interaction of RGB pixels within each colour plane and the interaction between different colour planes. Matas and Kittler [11] grouped colour pixels by taking into account simultaneously both their feature space similarity and spatial coherence.

None of the above approaches or any other colour segmentation work known to the authors have taken into consideration the interaction between colour and spatial frequency of patterns.

3   Building the Perceptual Tower

The resolution of an image signifies the area in physical units a pixel corresponds to. For example, 1 pixel $= 3 \times 3mm^2$ in the scene. When the same physical object is seen at a different distance, the resolution of the image changes, for example, 1 pixel $= 3 \times 3cm^2$. At the same time, the number of "pixels" the image of the object occupies in the retina reduces. Each pixel now carries the (blurred) information from several other pixels in the finer resolution version. Thus, when one blurs the image to imitate human vision, one should subsequently subsample the image as well. This way, a pyramid of image resolutions is created. We chose not to perform this subsampling, hence we create a tower of images instead of a pyramid. The reason is dual: (1) we like to keep the redundant information in the coarse levels to increase the robustness of the system, (2) we maintain a direct correspondence between the pixels across the resolution/scale levels. As we do not perform subsampling, the sizes of the blurring masks we use become larger and larger in number of pixels as we proceed to compute the coarser levels of the tower. Seen in this way, our approach is multiscale as filters of various scale sizes are employed. Therefore, throughout this paper, we do not distinguish between the terms multiresolution and multiscale.

The response characteristics of the human visual mechanism are functions of not only the spectral properties of the stimuli, but also of the temporal and spatial variations of these stimuli. When an observer deals with multi-coloured objects, with fine textures, their colour matching behaviour is affected by the spatial properties of the observed pattern. Furthermore, the human visual system will experience loss of detail at increasing distances away from the object. It perceives coloured textures at a large distance as areas of fairly uniform colour, whereas variations in luminance, e.g. at the borders between two textured areas, are still perceived. Therefore it is necessary to introduce a multiscale smoothing algorithm that smooths an image according to human perception. Zhang and Wandell [1] studied recently systematically the colour perception of human subjects for different frequencies of spatial colour variation. They proposed an algorithm for perceptual smoothing appropriate for evaluating image coding schemes. It is based on measurements in psychophysical studies which showed that discrimination and appearance of small-field or fine patterned colours differ from similar measurements made using large uniform fields. The human eye perceives high spatial frequencies of colour as a uniform colour instead of being able to separate these colours. An algorithm, which takes this into account must smooth the image in luminance and chrominance colour planes separately with different filter matrices for the planes. Zhang and Wandell[1] advocated the use of the opponent colour space, which consists of three different colour planes, O_1, O_2, O_3, representing the luminance, the red-green and the blue-yellow planes respectively. In the $O_1O_2O_3$ colour space, each of the planes is smoothed separately with two-dimensional spatial kernels, defined as sums of Gaussian functions with different values of σ. The result of this operation is that the luminance plane is blurred lightly, whereas the red-green and the blue-yellow planes are blurred more strongly. This spatial processing technique is pattern-colour separable. Zhang and Wandell's filtered representation was then transformed back to CIE-XYZ and then to CIE-Lab resulting in their Spatial CIE-Lab space, namely SCIE-Lab.

In this application, we set up three convolution matrices for the colour planes for each separate viewing distance. In any particular set, each of the three matrices consists of a weighted sum of Gaussian kernels. The matrices are computed according to[1]:

$$\frac{1}{m}\sum_i \frac{w_i}{n_i} e^{-\frac{x^2+y^2}{\sigma_i^2}} \qquad (1)$$

The values for (w_i, σ_i) which have been determined from psychological measurements of colour appearance on human subjects are given in [1] for a distance of 18 inches from the screen. We have derived new values for various distances by appropriate scaling. Divisor n_i in equation 1 is introduced to normalise the sum of the matrix elements of each individual Gaussian kernel before the weighted sum is applied. Divisor m normalises the sum of the final matrix to 1. These kernels are scaled, so that they each sum up to one.
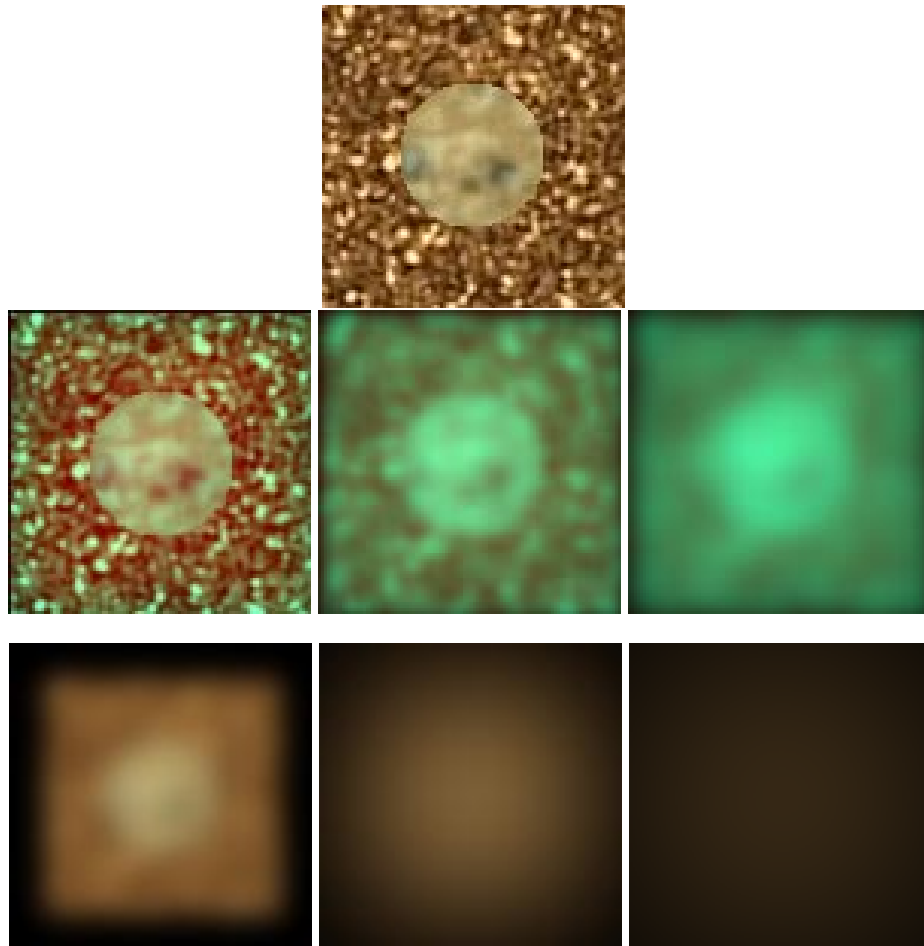
Fig. 1. (Row 1) Real texture collage image (Row 2) perceptual and (Row 3) Gaussian smoothed transformations corresponding to viewing distances of 1, 5, and 10 meters respectively.

Fig. 2. Real texture collage image with initial clusters and derived core clusters.

Once the kernels are applied to the image in the opponent colour space, we transform the image data from the $O_1O_2O_3$ to the CIE-Luv space and use it as input in the ensuing steps of the algorithm. This step is performed because the CIE-Luv space is a perceptually uniform space and therefore more suitable for carrying out colour measurements. Figure 1 shows a real texture collage and its associated smoothed images at varying distances for both perceptual and Gaussian smoothing. Clearly, the perceptually smoothed images provide a more realistic representation and blurring of an "object" viewed at varying distances. Most particularly, the Gaussian has vastly mixed and smoothed the colour values when convolved with each of the three colour channels.

4   Multiscale Probabilistic Relaxation

The problem of multiscale probabilistic labelling of the input image using a set of perceptually blurred versions of the image can be stated as follows. Let l indicate the levels of coarsening with $l = 1, ..., L$, representing the levels from full resolution to the coarsest level. Let $i, i = 1, ..., N$ be a pixel and $\mathbf{x}_i^l$, the associated measurement vector for that pixel at resolution level l. We define a label set Ω, $\Omega = \{\omega_1, \omega_2, ..., \omega_m\}$, which contains all possible labels of the image for m possible perceptual categories. Thus, each pixel i has label θ_i that can take on values from Ω.

We wish to choose for pixel i the most probable label θ_i given all the available information. In other words we wish to set:

$$\theta_i = arg\{\max_k P(\theta_i = \omega_k \mid \mathbf{x}_j^l, \forall j, \forall l)\} \tag{2}$$

For simplicity and clarity of exposition we shall restrict ourselves in considering only two successive levels of resolution l and $l+1$. Then, using Bayes's rule we have:

$$P(\theta_i = \omega_k \mid \mathbf{x}_j^l, \mathbf{x}_j^{l+1}, \forall j) = \frac{P(\theta_i = \omega_k, \mathbf{x}_j^l, \mathbf{x}_j^{l+1}, \forall j)}{P(\mathbf{x}_j^l, \mathbf{x}_j^{l+1}, \forall j)} \tag{3}$$

We can expand the terms in the numerator and the denominator by applying the theorem of total probability:

$$P(\theta_i = \omega_k \mid \mathbf{x}_j^l, \mathbf{x}_j^{l+1}, \forall j) =$$

$$\frac{\sum_{\omega_{\theta_1}} .. \sum_{\omega_{\theta_{i-1}}} \sum_{\omega_{\theta_{i+1}}} .. \sum_{\omega_{\theta_N}} P(\theta_1 = \omega_{\theta_1}, .., \theta_i = \omega_k, .., \theta_N = \omega_{\theta_N}, \mathbf{x}_j^l, \mathbf{x}_j^{l+1}, \forall j)}{\sum_{\omega_{\theta_1}} .. \sum_{\omega_{\theta_i}} .. \sum_{\omega_{\theta_N}} P(\theta_1 = \omega_{\theta_1}, .., \theta_i = \omega_{\theta_i}, .., \theta_N = \omega_{\theta_N}, \mathbf{x}_j^l, \mathbf{x}_j^{l+1}, \forall j)} \tag{4}$$

The joint probability that appears in equation 4 can be factorised as follows:

$$P(\theta_1 = \omega_{\theta_1}, ..., \theta_N = \omega_{\theta_N}, \mathbf{x}_j^l, \mathbf{x}_j^{l+1}, \forall j) =$$
$$P(\mathbf{x}_1^{l+1}, ..., \mathbf{x}_N^{l+1} \mid \theta_1 = \omega_{\theta_1}, ..., \theta_N = \omega_{\theta_N}, \mathbf{x}_1^l, ..., \mathbf{x}_N^l) \ \times$$
$$P(\theta_1 = \omega_{\theta_1}, ..., \theta_N = \omega_{\theta_N}, \mathbf{x}_1^l, ..., \mathbf{x}_N^l) \tag{5}$$

As we try to emulate here perceptual segmentation, we can imagine that due to causality, measurements obtained at level $l + 1$ (the coarser level) can not possibly depend on measurements obtained at level l. Thus, the first factor on the right hand side of equation 5 can be simplified as follows:

$$P(\mathbf{x}_1^{l+1}, ..., \mathbf{x}_N^{l+1} \mid \theta_1 = \omega_{\theta_1}, ..., \theta_N = \omega_{\theta_N}, \mathbf{x}_1^l, ..., \mathbf{x}_N^l) =$$
$$P(\mathbf{x}_1^{l+1}, ..., \mathbf{x}_N^{l+1} \mid \theta_1 = \omega_{\theta_1}, ..., \theta_N = \omega_{\theta_N}) \qquad (6)$$

We also expect that the measurement concerning a certain pixel depends on the identity of that pixel alone and on nothing else. Therefore, we can further write:

$$P(\mathbf{x}_1^{l+1}, ..., \mathbf{x}_N^{l+1} \mid \theta_1 = \omega_{\theta_1}, ..., \theta_N = \omega_{\theta_N}) =$$
$$\prod_j P(\mathbf{x}_j^{l+1} \mid \theta_j = \omega_{\theta_j}) = \prod_j \frac{P(\theta_j = \omega_{\theta_j} \mid \mathbf{x}_j^{l+1})}{\hat{p}(\theta_j = \omega_{\theta_j})} \, \hat{p}(\mathbf{x}_j^{l+1}) \qquad (7)$$

where $\hat{p}(\mathbf{x}_j^{l+1})$ is the prior probability of measurements $\mathbf{x}_j^{l+1}$ to arise, and $\hat{p}(\theta_j = \omega_{\theta_j})$ is the prior probability of label ω_{θ_j}. Now consider the second factor on the right hand side of equation 5:

$$P(\theta_1 = \omega_{\theta_1}, ..., \theta_N = \omega_{\theta_N}, \mathbf{x}_1^l, ..., \mathbf{x}_N^l) =$$
$$P(\mathbf{x}_1^l \mid \theta_1 = \omega_{\theta_1}, ..., \theta_N = \omega_{\theta_N}, \mathbf{x}_2^l, ..., \mathbf{x}_N^l) \ \times$$
$$P(\theta_1 = \omega_{\theta_1}, ..., \theta_N = \omega_{\theta_N}, \mathbf{x}_2^l, ..., \mathbf{x}_N^l) \qquad (8)$$

We can further expand the second term on the right hand side of equation 8 to write:

$$P(\theta_1 = \omega_{\theta_1}, ..., \theta_N = \omega_{\theta_N}, \mathbf{x}_1^l, ..., \mathbf{x}_N^l) =$$
$$P(\mathbf{x}_1^l \mid \theta_1 = \omega_{\theta_1}, ..., \theta_N = \omega_{\theta_N}, \mathbf{x}_2^l, ..., \mathbf{x}_N^l) \ \times$$
$$P(\mathbf{x}_2^l \mid \theta_1 = \omega_{\theta_1}, ..., \theta_N = \omega_{\theta_N}, \mathbf{x}_3^l, ..., \mathbf{x}_N^l) \ \times .... \times$$
$$P(\mathbf{x}_N^l \mid \theta_1 = \omega_{\theta_1}, ..., \theta_N = \omega_{\theta_N}) \ \times$$
$$P(\theta_1 = \omega_{\theta_1}, ..., \theta_N = \omega_{\theta_N}) \qquad (9)$$

For the same reasons explained earlier, we expect that the measurement obtained for a particular object depends on the identity of the object itself and on nothing else. Thus, all factors on the right hand side of equation 9, except the last one, can be simplified to express dependence only on the identity of the object they refer to. The last factor is the joint probability of a certain label assignment to arise. So we have:

$$P(\theta_1 = \omega_{\theta_1}, ..., \theta_N = \omega_{\theta_N}, \mathbf{x}_1^l, ..., \mathbf{x}_N^l) =$$
$$\prod_j P(\mathbf{x}_j^l \mid \theta_j = \omega_{\theta_j}) \ \times P(\theta_1 = \omega_{\theta_1}, ..., \theta_N = \omega_{\theta_N}) \qquad (10)$$

Now by substituting from equations 7 and 10 in equation 5 we obtain:

$$P(\theta_1 = \omega_{\theta_1}, ..., \theta_N = \omega_{\theta_N}, \mathbf{x}_j^l, \mathbf{x}_j^{l+1}, \forall j) =$$

$$\prod_j \frac{1}{\hat{p}(\theta_j = \omega_{\theta_j})} \, P(\theta_j = \omega_{\theta_j} \mid \mathbf{x}_j^{l+1}) \, \hat{p}(\mathbf{x}_j^{l+1}) \, P(\mathbf{x}_j^l \mid \theta_j = \omega_{\theta_j}) \times$$

$$P(\theta_1 = \omega_{\theta_1}, ..., \theta_N = \omega_{\theta_N}) \tag{11}$$

Then, upon substitution in equation 4:

$$P(\theta_i = \omega_k \mid \mathbf{x}_j^l, \mathbf{x}_j^{l+1}, \forall j) =$$

$$\frac{P(\mathbf{x}_i^l \mid \theta_i = \omega_k) \, P(\theta_i = \omega_k \mid \mathbf{x}_i^{l+1}) \, \hat{p}(\mathbf{x}_i^{l+1}) \prod_j \hat{p}(\mathbf{x}_j^{l+1}) \, Q(\theta_i = \omega_k)}{\sum_{\omega_{\theta_i}} P(\mathbf{x}_i^l \mid \theta_i = \omega_{\theta_i}) \, P(\theta_i = \omega_{\theta_i} \mid \mathbf{x}_i^{l+1}) \, \hat{p}(\mathbf{x}_i^{l+1}) \prod_j \hat{p}(\mathbf{x}_j^{l+1}) \, Q(\theta_i = \omega_{\theta_i})} \tag{12}$$

where

$$Q(\theta_i = \omega_{\theta_i}) =$$

$$\frac{1}{\hat{p}(\theta_i = \omega_{\theta_i})} \sum_{\omega_{\theta_1}} \cdots \sum_{\omega_{\theta_{i-1}}} \sum_{\omega_{\theta_{i+1}}} \cdots \sum_{\omega_{\theta_N}} \prod_j \frac{P(\mathbf{x}_j^l \mid \theta_j = \omega_{\theta_j}) P(\theta_j = \omega_{\theta_j} \mid \mathbf{x}_j^{l+1})}{\hat{p}(\theta_j = \omega_{\theta_j})} \times$$

$$P(\theta_1 = \omega_{\theta_1}, ..., \theta_N = \omega_{\theta_N}) \tag{13}$$

In the above expression $\hat{p}(\mathbf{x}_i^{l+1})$ is independent of the summation indices and cancels in the numerator and denominator. Therefore, equation 12 further simplifies to:

$$P(\theta_i = \omega_k \mid \mathbf{x}_j^l, \mathbf{x}_j^{l+1}, \forall j) = \frac{P(\mathbf{x}_i^l \mid \theta_i = \omega_k) \, P(\theta_i = \omega_k \mid \mathbf{x}_i^{l+1}) \, Q(\theta_i = \omega_k)}{\sum_{\omega_{\theta_i}} P(\mathbf{x}_i^l \mid \theta_i = \omega_{\theta_i}) \, P(\theta_i = \omega_{\theta_i} \mid \mathbf{x}_i^{l+1}) \, Q(\theta_i = \omega_{\theta_i})} \tag{14}$$

At the finest resolution equations 2 and 14 give the final labelling result.

5 Application to Colour texture Segments

In the previous two sections we presented the basic ingredients of our algorithm. The core of the multilevel probabilistic relaxation lies in the implementation of equations 13 and 14 and the estimation of the quantities that appear in them. The method works in a bootstrapping manner to estimate the various quantities needed. Thus, it is almost wholly unsupervised, with the possible exception of specifying the initial number of clusters in the coarsest level, if the K-means clustering method is used. It is possible to eliminate totally even this requirement by using a self-organising initial segmentation algorithm, like for example a watershed approach, or the method presented in [11], but we consider this as a point of secondary importance at present. In what follows we shall describe how each quantity that appears in 13 and 14 is estimated.

5.1 Core Clusters

The core clusters describe groups of pixels which can be confidently associated with the same region of texture in the image. The core clusters form the basis for setting up the colour histograms at different levels. To derive core clusters from the initial clusters, we need to fuzzify the segmentation/classification result obtained at the coarsest initialisation level. As this is only a step to help start the iteration process, we are adopting a rather simplistic approach: we first calculate the standard deviation σ_c of each cluster $c, c = 1, .., C$ where C is the total number of clusters. Then, we associate with every pixel a confidence, $\hat{p}_c^i$, with which it may be associated with each cluster:

$$\hat{p}_c^i \equiv \frac{\frac{\sigma_c^2}{d_c^{i\,2} + \sigma_c^2}}{\frac{\sigma_1^2}{d_1^{i\,2} + \sigma_1^2} + \frac{\sigma_2^2}{d_2^{i\,2} + \sigma_2^2} + ... + \frac{\sigma_C^2}{d_C^{i\,2} + \sigma_C^2}} \quad \forall i, \forall c \tag{15}$$

where $d_c^{i\,2}$ is the squared distance of pixel i from the mean of cluster c. Note that this formula has the property of giving a confidence higher than 50% to pixels that are closer than 1 σ from the mean of the cluster to belong to that cluster. Each core cluster is formed from the pixels that can be associated with it with a confidence of at least 80%. Figure 2 shows an example image with both its initial clusters and the subsequently derived core clusters.

Quantities $\hat{p}_c^i$ are also used to initialise the values of $P(\theta_i = \omega_{\theta_i} \mid \mathbf{x}_i^{l+1})$ which appear in 13 and 14. Thus, we set:

$$P(\theta_i = \omega_c \mid \mathbf{x}_i^L) = \hat{p}_c^i \quad \forall i, \forall c. \tag{16}$$

At all other levels $l < L$ these quantities are the probability label assignments computed for each pixel at level $l + 1$.

5.2 Prior Probabilities

The relative sizes of the core clusters are used as measures of the prior probabilities of the cluster labels, i.e. quantity $\hat{p}(\theta_j = \omega_{\theta_j})$ appearing in equation 13.

This is based on the observation that the larger clusters will appear most dominant when a texture mosaic is viewed from a large distance, and at the same time the prior probability of a pixel to belong to each cluster is proportional to the size of each cluster, in absence of any other information concerning the pixel.

5.3 3D Colour Histograms

The core clusters formed at resolution level $l + 1$ are mapped back into the image at resolution l and using the colour pixel values in those regions, a three dimensional colour histogram is set up (dynamically) for each region. This provides a statistical characterisation for each different texture at each resolution. From these colour histograms, the likelihood of a pixel i at smoothing stage l to have label ω_k can be calculated using the colour of this pixel. This likelihood is represented by $P(\mathbf{x}_i^l \mid \theta_i = \omega_k)$. Note that this way the distribution of the features that characterise a texture at each resolution level can be derived.

5.4 The Q-Function Pattern Dictionary

Equation 13 involves a summation over all possible labels of all pixels other than the pixel under consideration. Clearly, such a summation is impossible due to the enormous number of combinations one would have to consider. We prune the number of possibilities by imposing a limit to the number of pixels we shall consider as influencing the labelling of the pixel under consideration. Thus, instead of examining all other $N - 1$ pixels, we handle only a subset of them constituting a local neighbourhood around the pixel. We restrict this to be a 3×3 neighbourhood. This allows us then to introduce a dictionary of permissible label configurations within each 3×3 patch. As junctions are rare events in images, in most cases we have only 1 or at most 2 regions present in any 3×3 patch. Hence, we restrict the entries of our dictionary to be of the form presented in Figure 3 where A and B stand for any pair of cluster labels present in the image. All entries of the dictionary are assigned equal probability, thus factor $P(\theta_1 = \omega_{\theta_1}, ..., \theta_N = \omega_{\theta_N})$ in 13 becomes a constant and therefore redundant. Label combinations that do not appear in the dictionary have zero probability to exist and so they do not enter the summation on the right hand side of 13. Thus, equation 13 simplifies to

$$Q(\theta_i = \omega_{\theta_i}) = \frac{1}{\hat{p}(\theta_i = \omega_{\theta_i})} \sum_{\Re} \sum_{\Im} \prod_{\aleph} \frac{P(\mathbf{x}_j^l \mid \theta_j = \omega_{\theta_j}) P(\theta_j = \omega_{\theta_j} \mid \mathbf{x}_j^{l+1})}{\hat{p}(\theta_j = \omega_{\theta_j})} \quad (17)$$

where $\Re$ is the set of all patterns in the dictionary, $\Im$ is the set containing all possible combinations of two labels where centre pixel has label ω_{θ_i}, and $\aleph$ is the set of 3×3 pixel neighbourhood entries in dictionary

An improvement of the Q-function is possible by expanding the pattern dictionary to patterns with more than two different labels. It is also possible to calculate the Q-function for a neighbourhood larger than 3×3 pixels.
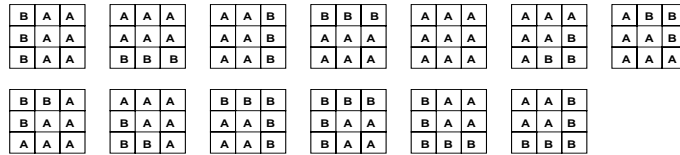

Fig. 3. Entries in pattern dictionary

6 Experimental Results

The results in this section are shown using arbitrarily selected colours to high-light different regions. All images tested were smoothed at different scales "or distances" (every meter up to $10m$) using the perceptual smoothing kernels. The processing commences from the distance at which the cluster histograms can be regarded as having separated modes. The results for the Gaussianly smoothed images were simply very wrong. To save space we do not report them here. Instead, we compare the results to an alternative approach (Matas and Kittler [11]) which is a more objective and comparable exercise.

The images shown on the left of Figure 4 are the original images of real texture collages put together from ceramic tile and granite stone textures. These textures are inherently random in nature. Our segmentation results are shown in the rightmost column while in the middle column the results from Matas and Kittler's [11] approach (hereafter referred to as MK) are demonstrated. MK exploit global and local image statistics simultaneously while also incorporating connectivity information. They discard the intensity information and form a 2D histogram of the chromaticity components. The image feature space is then partitioned by locating locally unimodal parts of the histogram. The spatial consistency of this segmentation is then examined and refined by incorporating neighbourhood connectivity. In the latter sense, our Q-function is also involving neighbourhood information. Moreover, we also consider more global contextual information by incorporating the prior label probabilities and the iterative re-finement of the initial segmentation. More poignantly, we encompass information from all three bands in our 3D histograms. Ignoring the intensity information can be useful if observing non-flat objects where the changes in intensity will deceive the observer's true colour perception. However, our purpose is to segment the scene as the observer views it, and not necessarily as the scene colours truly are. Therefore, considering the intensity band is very important.

As we have ground-truth information on these cases, the error measures for incorrectly classified pixels in both MK's and our perceptual segmentation are compared in Table 1. In every case we achieve a better segmentation. In the difficult test case T1, MK's technique finds a slightly smaller circle, therefore quite a number of pixels along the perimeter of the circle are misclassified. In test case T2, MK's technique has incorrectly combined the top-left and bottom-right patches as one class, while the latter is riddled with noisy segmentation; this is due to the non-unimodal representation of the guilty texture in MK's histograms, while it also is affected in a different way through its association with other types of texture in the image. In case T4, in which the image is a combination of real granite stone textures, the pixel values are quite spread out in the histogram, and there is little gradient there for MK's algorithm to work, while by perceptual segmentation the image can be correctly segmented (98.3%).

It is noteworthy that the perceptual segmentation algorithm has the ability to recover from incorrectly assigned pixels through the iterative relaxation process. Pixels already assigned to the wrong cluster, change their label again in future

Test Image	T1	T2	T3	T4
Matas & Kittler	2.2%	30%	1.3%	-
Perceptual	0.02%	1.2% %	0.04%	1.7%

Table 1. Error percentages of incorrectly classified pixels

steps thanks to the refinement in context, new neighbourhood information, and higher resolution information at consecutive levels.

The next set of results are more subjective and are expected to provide a more perception based representation of an image. Figures 5, 6, 7 respectively show the perceptual segmentation of a forest scene, the painting *La seine à argenteuil* by *Claude Monet*, and an aerial image. In all these cases, there are no nice straight edges that the Q-function could take advantage of to give an "accurate" segmentation. This is in fact the desired result as these images demonstrate the typical fuzzy segmentation of a scene that an observer may view from a distance.

Other important issues to note are that the histogram resolution in our experiments allows each bucket to cover an interval of 3.5 units in each direction in the Luv colour space. This is the resolution for the minimum perceivable colour distance for human vision. The clustering parameters are naturally very important since they determine the quality of the initial clusters on which our perceptual segmentation technique is based. However, we hope to use a parameter-free clustering approach in the future such as histogram watershed clustering[31].

The relaxation process can be iterated not only through the smoothed images, but also at each smoothed image for further incorporation and validation of image context. Naturally, this would add to the computational cost. At present, the smoothing stage demands a high computational cost due to the convolution filter sizes of Zhang and Wandell. However, the clustering and the relaxation process take approximately 60 seconds on a Silicon Graphics R10000 processor for a 128×128 image.

7 Discussions and Conclusions

Colour is an important parameter in the human visual experience. Most work in the past on texture analysis and segmentation has been concerned with deriving structural descriptors of texture, e.g. coarseness, regularity, blobiness, orientation etc., with the colour information perhaps used as an extra cue.

In this paper we treated the interplay of colours and their spatial distribution in an inseparable way as they are actually perceived during the pre-attentive stage of human colour vision. To do this we developed a tower of blurred versions of an image created by masks imitating the blurring the human vision sensors experience for scenes viewed at different distances, and allowed the information in this tower to flow in a causal direction, from the most blurred level to the most focussed. The creation of the tower made use of the latest results of psychophysics research, while the framework developed for the causal transfer of information

is quite general and can be applied for image segmentation where the features used could be other than colour.

Finally, the probabilistic relaxation methodology developed works in the opposite sense than other probabilistic relaxation schemes where the flow of information starts from the immediate neighbours of a pixel and, as the iteration steps progress, the influence of more distant pixels is incorporated through the succession of immediate neighbour interactions. In our case, probabilistic relaxation works in the same sense as all other multiresolution/multiscale schemes where first the information of long-range interaction is absorbed, followed by the information from the shorter range interaction. As we do not perform subsampling when we create the levels of the multiscale tower and we keep only the same immediate neighbours as contextual neighbourhood of a pixel, it may appear that we lack the mechanism to incorporate information from distant pixels. This is not so, because through the increasing size of the blurring masks we use to create the multiscale tower, the information from larger and larger distances is "smeared" into the immediate neighbours of a pixel and through interaction with them is incorporated into it.

References

1. X. Zhang and B.A. Wandell. A spatial extension of CIELAB for digital color image reproduction. In *Society for Information Display Symposium, San Diego*, 1996. WWW address: ftp://white.stanford.edu/scielab/spie97.ps.gz.

2. B.A. Wandell and X. Zhang. SCIELAB: a metric to predict the discriminability of colored patterns. In *9th Workshop on Image and Multidimensional Signal Processing*, pages 11–12, 1996.

3. S. Peleg, J. Naor, R. Hartley, and D. Avnir. Multiple resolution texture analysis and classification. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 6(4):518–523, 1984.

4. J.L. Crowley and A.C. Sanderson. Multiple resolution representation and probabilistic matching of 2-d gray-scale shape. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 9:113–120, 1987.

5. C. Bouman and B. Liu. Multiple resolution segmentation of textured images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(2):99–113, 1991.

6. S. Lam and H. Ip. Structural texture segmentation using irregular pyramid. *Pattern Recognition Letters*, 15:691–698, 1994.

7. F. Glazer. Multilevel relaxation in low-level computer vision. In A. Rosenfeld, editor, *Multiresolution Image Processing and Analysis*, pages 312–330. Springer, 1984.

8. D. Terzopoulos. Image analysis using multigrid relaxation methods. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(2):129–139, 1986.

9. D. Zhang, J. Liu, and F. Wan. Multiresolution relaxation: Experiments and evaluations. In *Proceedings of International Conference on Pattern Recognition*, pages 712–714, 1988.

10. E. R. Hancock, M. Haindl, and J. Kittler. Multiresolution edge labelling using hierarchical relaxation. In *Proceedings of International Conference on Pattern Recognition*, pages 140–144, 1992.

11. J. Matas and J. Kittler. Spatial and feature based clustering: Applications in image analysis. In *CAIP95*, pages 162–173, 1995.
12. B. Julesz and J.R. Bergen. Textons, the fundamental elements in preattentive vision and perception of textures. *Bell Systems Technical Journal*, 62(6):1619–1645, 1983.
13. J. Malik and P. Perona. A computational model of texture perception. Technical Report CSD-89-491, University of California Berkeley, CS, 1989.
14. H. Tamura, S. Mori, and T. Yamawaki. Textural features corresponding to visual perception. *SMC*, 8(6):460–473, June 1978.
15. R. M. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804, 1979.
16. L. V. Gool, P. Dewaele, and A. Oosterlinck. Texture analysis anno 1983. *Computer Vision, Graphics and Image Processing*, 29:336–357, 1985.
17. T.R. Reed and J. du Buf. A review of recent texture segmentation and feature extraction techniques. *CVGIP: Image Understanding*, 57:359–372, 1993.
18. A.K. Jain and F. Farrokhnia. Unsupervised texture segmentation using gabor filters. *Pattern Recognition*, 24(12):1167–1186, 1991.
19. M. Unser and M. Eden. Multiresolution feature extraction and selection for texture segmentation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11(7):717–728, July 1989.
20. I. Matalas, S. Roberts, and H. Hatzakis. A set of multiresolution texture features suitable for unsupervised image segmentation. In *Proceedings of Signal Processing VIII, Theories and Applications*, volume III, pages 1495–1498, 1996.
21. S.J. Roan, J.K. Aggarwal, and W.N. Martin. Multiple resolution imagery and texture analysis. *Pattern Recognition*, 20(1):17–31, 1987.
22. Y. Ohta, T. Kanade, and T. Sakai. Color information for region segmentation. *Computer Graphics and Image Processing*, 13:222–241, 1980.
23. G. Healey. Segmenting images using normalized color. *IEEE Trans. Systems, Man, and Cybernetics*, 22(1):64–73, 1992.
24. J. Liu and Y-H. Yang. Multiresolution color image segmentation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16:689–700, July 1994.
25. W. Skarbek and A. Koschan. Colour image segmentation - a survey. Technical report, Technical University Berlin, 1994.
26. M.J. Swain. *Color Indexing*. PhD thesis, University of Rochester, 1990.
27. G. J. Klinker. *A Physical Approach to Color Image Understanding*. A K Peters, Wellesley, Massachusetss, 1993.
28. G. B. Coleman and H. C. Andrews. Image segmentation by clustering. *Proceedings of the IEEE*, 67(5):773–785, 1979.
29. S. C. Tan and J. Kittler. Colour texture classification using features from colour histogram. *Proceedings of the Eighth Scandinavian Conference on Image Processing*, 1993.
30. D.K. Panjwani and G. Healey. Unsupervised segmentation of textured color images using markov random field models. In *Conference on Computer Vision and Pattern Recognition*, pages 776–777, 1993.
31. L. Shafarenko, M. Petrou, and J. Kittler. Automatic watershed segmentation of randomly textured color images. *IEEE Transactions on Image Processing*, 6(11):1530–1544, November 1997.
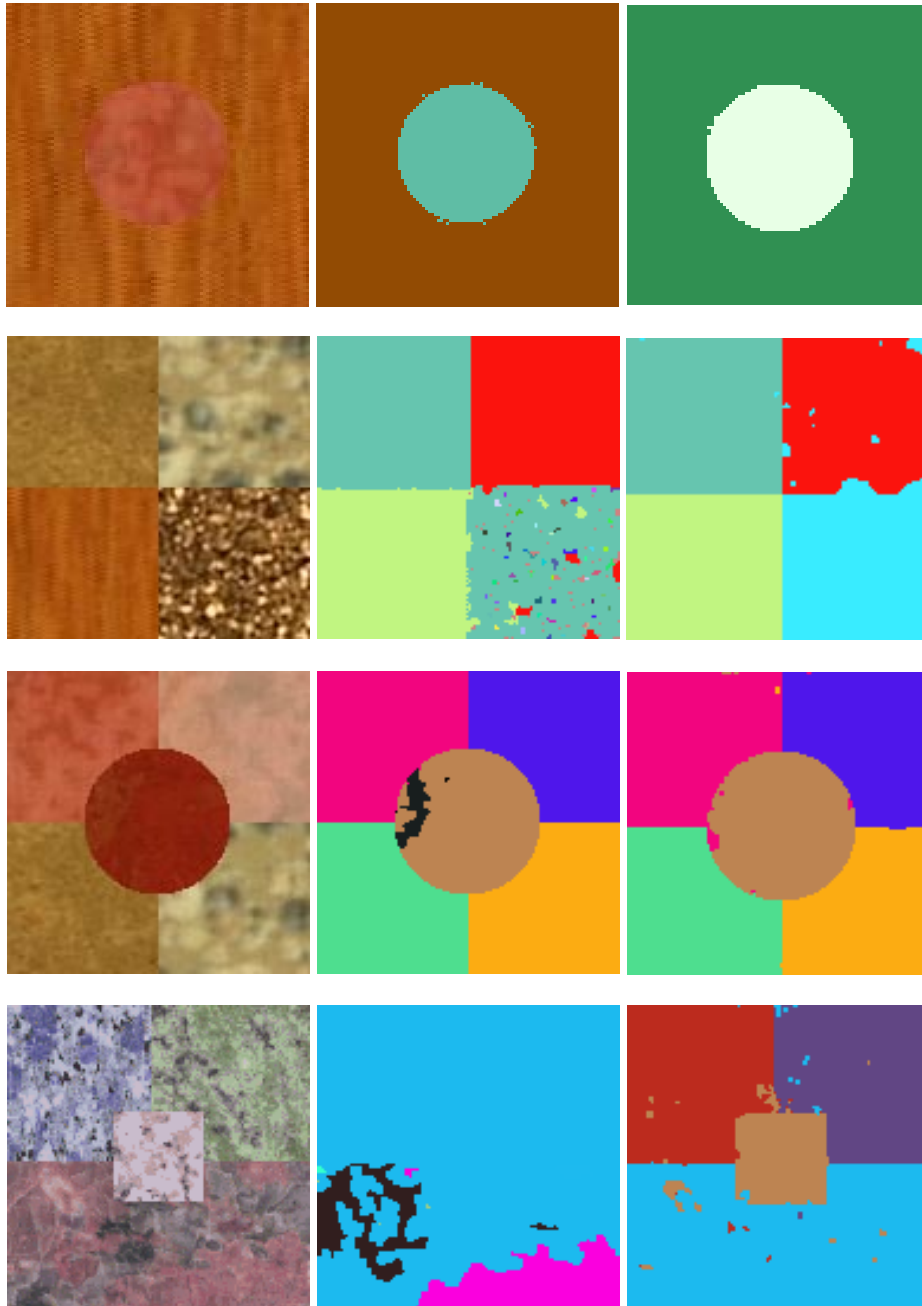
Fig. 4. Four test cases (T1-T4) and their segmentation by applying (mid-column) Matas & Kittler's approach and (right-column) perceptual segmentation.
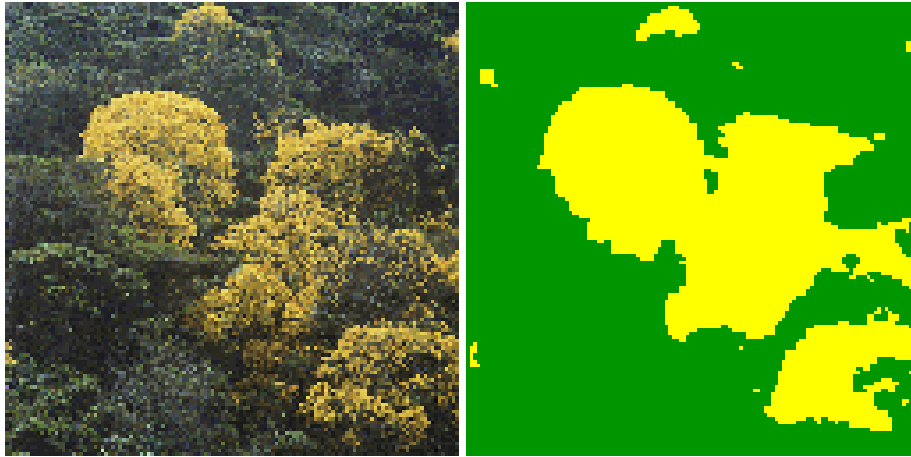
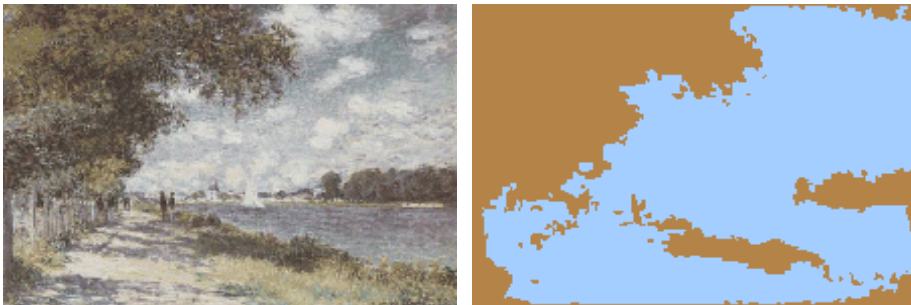**Fig. 5.** Forest image and its perceptual segmentation.
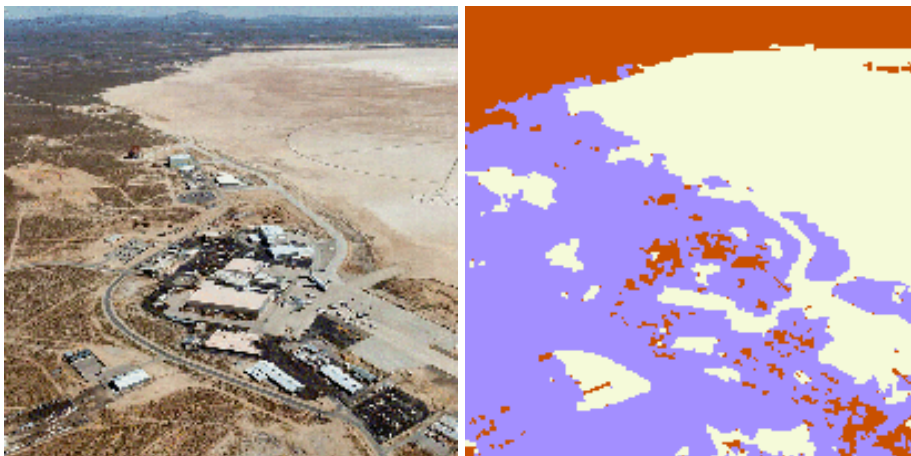

Fig. 6. Monet's painting and its perceptual segmentation.

Fig. 7. Land image and its perceptual segmentation.